Organic Chemistry

With a Biological Emphasis

Volume I: Chapters 1-9

Tim Soderberg
University of Minnesota, Morris

February 2012

Notes to the reader:

- This textbook is meant to be a constantly evolving work in progress, and as such, feedback from students, instructors, and all other readers is greatly appreciated. Please send any comments, suggestions, or notification of errors to the author at soderbt@morris.umn.edu.

- If you are looking at a printed version of this textbook, please be aware that most of the figures throughout contain color. Although they are intended to be understandable in black and white, it may often be very helpful to download (for free!) a pdf version of the text at lulu.com in order to refer to the color figures.

- An online version of this textbook is accessible as part of the **Chemwiki** project at the University of California, Davis: http://chemwiki.ucdavis.edu/Organic_Chemistry/Organic_Chemistry_With_a_Biological_Emphasis

- *Where is the index?* There is no printed index for this textbook. However, an electronic index is available simply by opening the digital (pdf) version of the text (download it for free at lulu.com) and using the 'find' or 'search' function of your pdf viewer.

Table of Contents

Volume I: Chapters 1-9

Introduction

Chapter 1: Introduction to organic structure and bonding I

Introduction
Section 1: Atomic orbitals and electron configuration
  A: The atom
  B: Atomic orbitals
  C: Electron configuration
Section 2: Chemical Bonds
  A: Ionic bonds
  B: Covalent bonds and Lewis structures
  C: Formal charges
Section 3: Drawing organic structures
  A: Common bonding patterns in organic structures
  B: Using the 'line structure' convention
  C: Constitutional isomers
  D: The Index of Hydrogen Deficiency
Section 4: Functional groups and organic nomenclature
  A: Common functional groups in organic compounds
  B: Naming organic compounds
  C: Abbreviated organic structures
Section 5: Valence bond theory
  A: Formation of sigma bonds: the H_2 molecule
  B: Hybrid orbitals: sp^3 hybridization and tetrahedral bonding
  C: Formation of pi bonds: sp^2 and sp hybridization
  D: The valence bonding picture in carbocations, carbanions, and carbon free radicals

Chapter 2: Introduction to organic structure and bonding II

Introduction
Section 1: Molecular orbital theory
  A: Another look at the H_2 molecule: bonding and antibonding sigma molecular orbitals
  B: MO theory and pi bonds: conjugation
  C: Aromaticity
Section 2: Resonance
  A: The meaning of resonance contributors: benzene and its derivatives
  B: Resonance contributors of the carboxylate group
  C: Rules for drawing resonance structures

- D: Major vs minor resonance contributors - four more rules to follow
- E: More examples of resonance: peptide bonds, enolates, and carbocations

Section 3: Non-covalent interactions
- A: Dipoles
- B: Ion-ion, dipole-dipole and ion-dipole interactions
- C: van der Waals forces
- D: Hydrogen bonds

Section 4: The relationship between noncovalent interactions physical properties
- A: Solubility
- B: Illustrations of solubility concepts - metabolic intermediates, lipid bilayer membranes, soaps and detergent
- C: Boiling points and melting points
- D: The melting behavior of lipid structures

Chapter 3: Conformations and Stereochemistry

Introduction

Section 1: Conformations of straight-chain organic molecules
- A: Conformations of ethane
- B: Conformations of butane

Section 2: Conformations of cyclic organic molecules
- A: Introduction to sugars and other cyclic molecules
- B: Ring size
- C: Conformations of glucose and other six-membered ring structures
- D: Conformations of pentose and other five-membered ring structures
- E: The importance of conformation in organic reactivity

Section 3: Stereoisomerism – chirality, stereocenters, enantiomers

Section 4: Defining stereochemical configuration - the Cahn-Ingold-Prelog system

Section 5: Interactions between chiral molecules and proteins

Section 6: Optical activity

Section 7: Diastereomers
- A: Compounds with multiple stereocenters
- B: *Meso* compounds
- C: Stereoisomerism of alkenes

Section 8: Fischer and Haworth projections

Section 9: Stereochemistry and organic reactivity

Section 10: Prochirality
- A: Prochiral substituents on tetrahedral carbons
- B: Carbonyl and imine carbons as prochiral centers

Chapter 4: Structure determination part I: Infrared spectroscopy, UV-visible spectroscopy, and mass spectrometry

Introduction

Section 1: Introduction to molecular spectroscopy

  - A: The electromagnetic spectrum
  - B: Molecular spectroscopy – the basic idea
- Section 2: Infrared spectroscopy
- Section 3: Ultraviolet and visible spectroscopy
  - A: Electronic transitions
  - B: Looking at UV-vis spectra
  - C: Applications of UV spectroscopy in organic and biological chemistry
- Section 4: Mass Spectrometry
  - A: The basics of mass spectrometry
  - B: Looking at mass spectra
  - C: Gas Chromatography - Mass Spectrometry
  - D: Mass spectrometry of proteins - applications in proteomics

Chapter 5: Structure determination part II- Nuclear magnetic resonance spectroscopy

- Introduction
- Section 1: The origin of the NMR signal
  - A: NMR-active nuclei
  - B: Nuclear precession, spin states, and the resonance condition
- Section 2: Chemical equivalence
- Section 3: The NMR experiment
  - A: The basics of an NMR experiment
  - B: The chemical shift
  - C: Signal integration
- Section 4: The basis for differences in chemical shift
  - A: Diamagnetic shielding and deshielding
  - B: Diamagnetic anisotropy
  - C: Hydrogen-bonded protons
- Section 5: Spin-spin coupling
  - A: The source of spin-spin coupling
  - B: Coupling constants
  - C: Complex coupling
- Section 6: ^{13}C-NMR spectroscopy
  - A: The basics of ^{13}C-NMR spectroscopy
  - B: ^{13}C-NMR in isotopic labeling studies
- Section 7 : Determining unknown structures
- Section 8: NMR of phosphorylated molecules

Chapter 6: Introduction to organic reactivity and catalysis

- Introduction
- Section 1: A first look at reaction mechanisms
  - A: An acid-base (proton transfer) reaction
  - B: A one-step nucleophilic substitution reaction (S_N2)
  - C: A two-step nucleophilic substitution reaction (S_N1)
- Section 2: Describing the thermodynamics and kinetics of chemical reactions - energy diagrams

Section 3: Enzymatic catalysis - the basic ideas
Section 4: Protein structure
    A: Amino acids and peptide bonds
    B: Visualizing protein structure: X-ray crystallography
    C: The four levels of protein structure
    D: The molecular forces that hold proteins together
Section 5: How enzymes work
    A: The active site
    B: Transition state stabilization
    C: Site-directed mutagenesis
    D: Enzyme inhibition
    E: Catalysts in the laboratory

Chapter 7: Organic compounds as acids and bases

Introduction
Section 1: The 'basic' idea of an acid-base reaction
    A: The Brønsted-Lowry definition of acidity
    B: The Lewis definition of acidity
Section 2: Comparing the acidity and basicity of organic functional groups– the acidity constant
    A: Defining K_a and pK_a
    B: Using pKa values to predict reaction equilibria
    C: pK_a and pH: the Henderson-Hasselbalch equation
Section 3: Structural effects on acidity and basicity
    A: Periodic trends
    B: The resonance effect
    C: The inductive effect
Section 4: More on resonance effects on acidity and basicity
    A: The acidity of phenols
    B: The basicity of nitrogen-containing groups: aniline, imines, pyridine, and pyrrole
Section 5: Carbon acids and enolate ions
Section 6: Polyprotic acids
Section 7: The effects of solvent and enzyme microenvironment on acidity

Chapter 8: Nucleophilic substitution reactions part I

Introduction
Section 1: Introduction to the nucleophilic substitution reaction
Section 2: Two mechanistic models for a nucleophilic substitution reaction
    A: Associative nucleophilic substitution: the S_N2 reaction
    B: Dissociative nucleophilic substitution: the S_N1 reaction
    C: Nucleophilic substitutions occur at sp^3-hybridized carbons
Section 3: More about nucleophiles
    A: What makes a nucleophile?
    B: Protonation states and nucleophilicity
    C: Periodic trends and solvent effects in nucleophilicity

    - D: Resonance effects on nucleophilicity
    - E: Steric effects on nucleophilicity
- Section 4: Electrophiles and carbocation stability
    - A: Steric effects on electrophilicity
    - B: Stability of carbocation intermediates
- Section 5: Leaving groups
    - A: What makes a good leaving group?
    - B: Leaving groups in biochemical reactions
    - C: Synthetic parallel - conversion of alcohols to alkyl halides, tosylates and mesylates
    - D: S_N1 or S_N2? Predicting the mechanism
- Section 6: Epoxides as electrophiles in nucleophilic substitution reactions
    - A: Epoxide structure
    - B: Epoxide ring-opening reactions - S_N1 vs S_N2, regioselectivity, and stereoselectivity

Chapter 9: Nucleophilic substitution reactions part II

- Introduction
- Section 1: Methyl group transfers: examples of S_N2 reactions
    - A: SAM methyltransferase
    - B: Synthetic parallel – the Williamson ether synthesis
- Section 2: Digestion of carbohydrate by glycosidases - an S_N1 reaction
- Section 3: Protein prenyltransferase - a hybrid S_N1/S_N2 substitution
    - A: The biological relevance of the protein prenyltransferase reaction
    - B: Determining the mechanism of protein prenyltransferase with fluorinated substrate analogs
    - C: The zinc-thiolate interaction in protein prenyltransferase - 'tuning' the nucleophile
- Section 4: Biochemical nucleophilic substitutions with epoxide electrophiles
    - A: Hydrolysis of stearic acid epoxide: investigating the mechanism with kinetic isotope effect experiments
    - B: Fosfomycin - an epoxide antibiotic
- Section 5: Nucleophilic substitution over conjugated pi systems - the SN' mechanism

Tables

- Table 1: Some characteristic absorption frequencies in IR spectroscopy
- Table 2: Typical values for ^{1}H-NMR chemical shifts
- Table 3: Typical values for ^{13}C-NMR chemical shifts
- Table 4: Typical coupling constants in NMR
- Table 5: The 20 common amino acids
- Table 6: Structures of common coenzymes
- Table 7: Representative acid constants
- Table 8: Some common laboratory solvents, acids, and bases
- Table 9: Examples of common functional groups in organic chemistry

Volume II: Chapters 10-17

Chapter 10: Phosphoryl transfer reactions

Introduction: Abundance of phosphoryl groups in metabolic intermediates
Section 1: Overview of phosphates and phosphoryl transfer reactions
    A: Nomenclature and abbreviations
    B: Acid constants and protonation states
    C: Bonding in phosphines and phosphates
    D: Phosphoryl transfer reactions - the general picture
    E: Phosphoryl transfer reactions - associative, addition-elimination, or dissociative?
Section 2: Phosphorylation reactions - kinase enzymes
    A: ATP - the principle phosphoryl group donor
    B: Monophosphorylation of alcohols
    C: Diphosphorylation of alcohols
    D: Phosphorylation of carboxylates
    E: Generation of nucleotide phosphates
    F: Regeneration of ATP from ADP
Section 3: Hydrolysis of phosphates
Section 4: Phosphate diesters
    A: Phosphate diesters as the backbone for DNA and RNA
    B: The chemistry of genetic engineering

Chapter 11: Nucleophilic carbonyl addition reactions

Introduction
Section 1: Nucleophilic additions to aldehydes and ketones: the general picture
Section 2: Stereochemistry of the nucleophilic addition reaction
Section 3: Hemiacetals, hemiketals, and hydrates
    A: The general picture
    B: Simple sugars are hemiacetals and hemiketals
Section 4: Acetals and ketals
    A: Glycosidic bonds revisited
    B: Synthetic parallel: cyclic acetals/ketals as 'protecting groups' for ketones and aldehydes
Section 5: N-glycosidic bonds
Section 6: Imine (Schiff base) formation
    A: Imines-the general picture
    B: Pyridoxal phosphate coenzyme links to enzymes by a Schiff base
    C: Schiff base formation in aldolase reactions
Section 7: A look ahead: addition of carbon and hydride nucleophiles to carbonyls

Chapter 12: Acyl substitution reactions

Introduction
Section 1: Introduction to carboxylic acid derivatives and the nucleophilic acyl substitution reaction
    A: Carboxylic acid derivatives and acyl groups
    B: The nucleophilic acyl substitution reaction
    C: The relative reactivity of carboxylic acid derivatives
Section 2: Acyl phosphates as activated carboxylic acids
    A: Glutamine synthetase
    B: Asparagine synthetase
    C: Glycinamide ribonucleotide synthetase
    12.2D: Synthetic parallel - activated carboxylic acids in the lab
Section 3: Thioesters
    A: Introduction to thioesters and Coenzyme A
    B: Activation of fatty acids by coenzyme A - a thioesterification reaction
    C: Transfer of fatty acyl groups to glycerol: a thioester to ester substitution
    D: More transthioesterification reactions
    E: Hydrolysis of thioesters
Section 4: Esters
    A: Nonenzymatic esterification: synthesis of 'banana oil'
    B: Nonenzymatic ester hydrolysis and the soap-making process
    C: Enzymatic ester hydrolysis: acetylcholinesterase and sarin nerve gas
    D: More enzymatic ester hydrolysis: lipase, the resolution of enantiomers, and dehalogenation
    E: Transesterification: the chemistry of aspirin and biodeisel
Section 5: Nucleophilic acyl substitution reactions involving peptide bonds
    A: Formation of peptide bonds on the ribosome
    B: Hydrolysis of peptide bonds: HIV protease
    C: The chemical mechanism of penicillin
Section 6: Activated amide groups
Section 7: A look ahead: acyl substitution reactions with a carbon or hydride nucleophile

Chapter 13: Reactions with stabilized carbanion intermediates, part I - isomerization, aldol and Claisen condensation, and decarboxylation

Introduction
Section 1: Tautomers
    A: Keto-enol tautomerization
    B: Imine/enamine tautomerization
Section 2: Isomerization reactions
    A: Carbonyl isomerization
    B: Stereoisomerization at chiral carbons
    C: Alkene isomerization in the degradation of unsaturated fatty acids
Section 3: Aldol reactions
    A: The general mechanism for an aldol reaction

  - B: Typical aldolase reactions: three variations on a theme
  - C: Going backwards: the retroaldol reaction
  - D: Going both ways: transaldolase
- Section 4: Claisen reactions
  - A: Claisen condensations
  - B: Retro-Claisen cleavages
  - C: Enolates as nucleophiles in S_N2 displacements
- Section 5: Carboxylation and decarboxylation reactions
  - A: The metabolic context of carboxylation and decarboxylation
  - B: The carboxylation mechanism of Rubisco
  - C: Decarboxylation
  - D: Biotin is a CO_2-carrying coenzyme
- Section 6: Synthetic parallel - carbon nucleophiles in the lab
  - A: Lab reactions with enolate /enamine intermediates
  - B: The Wittig reaction
  - C: Terminal alkynes as carbon nucleophiles
  - D: Grignard, Gilman, and organolithium reagents

Chapter 14: Reactions with stabilized carbanion intermediates, part II: Michael additions, eliminations, and electron sink cofactors

- Introduction
- Section 1: Michael additions and β-eliminations
  - A: Overview of Michael addition and β-elimination mechanisms
  - B: Stereochemistry of Michael additions and β-eliminations
  - C: NMR experiments to determine the stereochemistry of a Michael addition
  - D: More examples of elimination and addition reactions
- Section 2: Variations on the Michael reaction
  - A: *Cis/trans* alkene isomerization
  - B: Nucleophilic aromatic substitution
  - C: Synthetic parallel - Michael addition reactions in the laboratory
- Section 3: Elimination by the E1 and E2 mechanisms
  - A: E1 and E2 reactions in the laboratory
  - B: Enzymatic E1 and E2 reactions
- Section 4: Pyridoxal phosphate - an electron sink cofactor
  - A: PLP and the Schiff-base linkage to lysine
  - B: PLP-dependent amino acid racemases
  - C: PLP-dependent decarboxylation
  - D: PLP-dependent retroaldol reactions
  - E: PLP-dependent transaminase reactions (aspartate aminotransferase)
  - F: PLP-dependent β-elimination and β-substitution reactions
  - G: PLP-dependent γ-elimination and γ-substitution reactions
  - H: Altering the course of a PLP reaction through site-directed mutagenesis
- Section 5: Thiamine diphosphate-dependent reactions
  - A: The benzoin condensation reaction
  - B: The transketolase reaction

  C: Pyruvate decarboxylase
  D: Synthetic parallel - carbonyl nucleophiles *via* dithiane anions
Section 6: The transition state geometry of reactions involving pi bonds
  A: Transition state geometry of E2 reactions
  B: Transition state geometry of PLP-dependent reactions

Chapter 15: π electrons as nucleophiles: electrophilic additions, addition/eliminations, and rearrangements

Introduction
Section 1: An overview of the different types of electrophilic reactions
Section 2: Electrophilic addition
  A: The general picture
  B: The regiochemistry of electrophilic addition
  C: Enzymatic electrophilic additions
  D: Synthetic parallel - electrophilic additions in the laboratory
Section 3: Electrophilic isomerization and substitution (addition/elimination)
  A: Alkene isomerization
  B: Substitution by electrophilic addition/elimination
Section 4: Another kind of electrophilic addition-elimination - Shikimate to chorismate
Section 5: Electrophilic aromatic substitution
  A: The general picture
  B: Some representative enzymatic electrophilic aromatic substitution reactions
Section 6: Synthetic parallel - electrophilic aromatic substitution in the lab
  A: Friedel-Crafts reactions
  B: Ring directing effects in S_EAr reactions
Section 7: Carbocation rearrangements
  A: Hydride and alkyl shifts
  B: Enzymatic reactions with carbocation rearrangement steps
  C: The acyloin, pinacol, and Hoffman rearrangements (*isoleucine biosynthesis*).
Section 8: Cation-pi interactions and the stabilization of carbocation intermediates
Section 9: Outside the box - 1,3-elimination and rearrangement in squalene synthase
Section 10: The Diels-Alder reaction and other pericyclic reactions

Chapter 16: Oxidation and reduction reactions

Introduction
Section 1: Oxidation and reduction of organic compounds - an overview
Section 2: The importance of redox reactions in metabolism
Section 3: Outside the box - methanogenesis
Section 4: Hydrogenation/dehydrogenation reactions of carbonyls, imines, and alcohols
  A: Nicotinamide adenine dinucleotide - a hydride transfer coenzyme
  B: Carbonyl hydrogenation and alcohol dehydrogenation - the general picture
  C: Stereochemistry of carbonyl hydrogenation and alcohol dehydrogenation
  D: Examples of redox reactions involving alcohols, carbonyl groups, and imines
Section 5: Hydrogenation of alkenes and dehydrogenation of alkanes

  - A: Alkene hydrogenation in fatty acid biosynthesis
  - B: The flavin coenzymes
  - C: Alkane dehydrogenation in fatty acid degradation
  - D: More examples of enzymatic alkene hydrogenation
- Section 6: Additional examples of enzymatic hydride transfer reactions
  - A: More reactions involving nicotinamide adenine dinucleotide and flavin
  - B: Reactions with coenzymes derived from folic acid
- Section 7: NAD(P)H, $FADH_2$ and metabolism - a second look
  - A: NADH and $FADH_2$ as carriers of hydrides from fuel molecules to water
  - B: The source of NADPH for reductive biosynthesis
- Section 8: Observing the progress of hydrogenation and dehydrogenation reactions by UV assay
- Section 9: Hydrogenation/dehydrogenation reactions and renewable energy technology
- Section 10: Oxygenase reactions- flavin-dependent monoxygenases
- Section 11: Halogenation of organic compounds
  - A: Enzymatic halogenation
  - B: Synthetic parallel - halogenation of alkenes in the lab
- Section 12: Redox reactions involving thiols and disulfides
  - A: Disulfide bridges in proteins
  - B: The role of disulfides in the pyruvate dehydrogenase reaction
- Section 13: Redox reactions in the organic synthesis laboratory
  - A: Metal hydride reducing agents
  - B: Catalytic hydrogenation and the *trans* fat issue
  - C: Reduction of carbonyl carbons to methylene
  - D: Laboratory oxidation reactions

Chapter 17: Radical reactions

- Introduction
- Section 1: Structure and reactivity of radical species
  - A: The geometry and relative stability of carbon radicals.
  - B: The diradical character of triplet oxygen
- Section 2: Radical chain reactions
  - A: The three phases of radical chain reactions
  - B: Radical halogenation in the lab
  - C: Useful polymers formed by radical chain reactions
  - D: Destruction of the ozone layer by CFC radicals
  - E: Harmful radical species in cells and natural antioxidants
- Section 3: Enzymatic reactions with free radical intermediates
  - A: Hydroxylation of alkanes
  - B: Reductive dehydroxylation of alcohols
  - C: Radical mechanisms for flavin-dependent reactions

Tables

Table 1: Some characteristic absorption frequencies in IR spectroscopy
Table 2: Typical values for ^{1}H-NMR chemical shifts
Table 3: Typical values for ^{13}C-NMR chemical shifts
Table 4: Typical coupling constants in NMR
Table 5: The 20 common amino acids
Table 6: Structures of common coenzymes
Table 7: Representative acid constants
Table 8: Some common laboratory solvents, acids, and bases
Table 9: Examples of common functional groups in organic chemistry

Appendix I: Enzymatic reactions by metabolic pathway and EC number
Appendix II : Review of laboratory synthesis reactions

Introduction

As you begin a course in organic chemistry, you probably have many questions. Here are my attempts to answer a few of the questions that you are most likely to be wondering about.

What is organic chemistry, and why do I need to study it?

As you probably already know, organic chemistry is defined as the study of molecules that contain the element carbon. If you are interested in the science of living things, then you are also interested in organic chemistry - organic molecules and the reactions they undergo form the basic currency of life on earth. You cannot understand how a human cell breaks down carbohydrate or fat for energy without gaining a solid understanding of the basic processes of organic chemistry that underlie these metabolic events. You cannot appreciate how drugs work, on a molecular level, until you have first learned about the three dimensional structure of organic molecules and how different organic 'functional groups' interact with one another. And if you don't know your organic chemistry, you will have difficulty understanding and explaining the molecular underpinnings of diseases such as depression, cancer, or diabetes.

The relevance of organic chemistry is not restricted to the study of living systems, however. You need a firm grasp of the subject to understand many ongoing developments in renewable energy, nanomaterials, environmental clean-up, and drug development and regulation, just to name a few examples. Society always has - and probably always will - need people who know how to synthesize, analyze, and break down organic molecules in useful ways, and this need translates to rewarding careers with good salaries and working conditions.

How is organic chemistry going to be different from general chemistry?

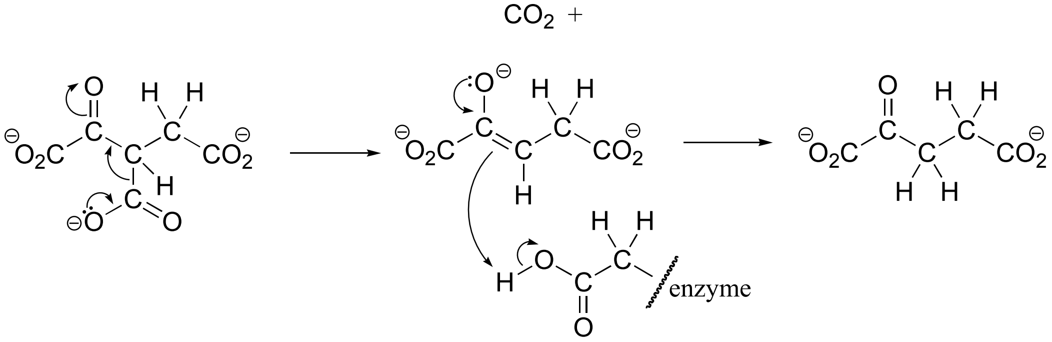
Depending on your experience with general chemistry, you may be pleased (or dismayed) to learn that most students find organic chemistry to be an entirely new breed of beast. It is common to see students, who felt very much at home with general chemistry, struggle with organic – and vice versa.

To begin with, an introductory course in organic chemistry tends to focus much more on the qualitative than it is does on the quantitative (translation – there much less math!) But don't get too comfortable just yet – despite its scarcity of math, you undoubtedly are aware that very few people find organic to be an easy course. In fact, many science and pre-health majors report that it was for them the single most challenging part of their undergraduate career. You are going to be asked to take in and digest a lot of new concepts, many of which involve visualization in three dimensions, and some of which are, at first pass at least, quite abstract. But that is just a start – the really hard part comes when you are asked to connect several of the ideas that you have learned, in a process of deductive reasoning. It is not enough, in other words, to understand A, B, and C as isolated concepts– you must also be able to solve chemical problems for which you need to find and follow the connections from A to B to C in order to arrive at a valid solution. It is this aspect of organic chemistry that students find the most challenging, yet also the most rewarding.

Another major difference between general chemistry and organic chemistry has to do with the way you will be thinking about chemical reactivity. In general chemistry, you mainly saw reactions depicted by chemical equations of the type:

$$C_6H_3O_7 \longrightarrow CO_2(g) + C_5H_4O_5$$

In this type of treatment, you focused mainly on changes in molecular formulas, and on basic ideas of thermodymanics (energy changes) and kinetics (rates of reactions). For the most part, you disregarded the structure of the molecules involved and the details of how the reaction proceeded. In organic chemistry, we will be very concerned with the structures of the reacting molecules, and we will also think very carefully about how the reaction takes place – which bonds break, which bonds form, and in what order this happens.



This is known as the **mechanism** of a reaction.

We will also think carefully about why reactions take place as they do: we'll be thinking, in other words, about the connections between structure and reactivity, and also about how chemical structure relates to the thermodynamics and kinetics of an equation. Why, for example, does one specific carbon-carbon bond break in the reaction sequence above, and not others? What is the importance of the carbon-oxygen double bonds? How is the reaction accelerated by an enzyme catalyst?

By necessity, an understanding of structure must come before we tackle reactivity. We will spend the first several chapters learning about how molecules are put together (their structure), and then turn, for the remainder of the book, to discussion about how our knowledge of molecular structure can help us to understand the many different ways in which organic molecules react with each other. By the end of the course, your understanding of chemical reactions will have taken on entirely new dimensions.

How is this textbook different from other organic chemistry textbooks?

In this text, we focus our attention primarily on biological organic chemistry - the organic reactions that take place in living things - simply because this is, for most people, the most interesting and relevant context in which to learn the principles of the subject. This is not, however, the traditional way to approach the subject – if you take a look at most other organic chemistry textbooks, you will see a much stronger emphasis on examples that come not from biology but from the chemical synthesis laboratory, where chemists work to create new synthetic molecules for medicinal or industrial use. In this text we will study some of the most important reactions of laboratory synthesis with which professional chemists need to be familiar - these are usually in sections entitled 'synthetic parallel' in order to highlight the conceptual similarities between laboratory and biological chemistry. You will notice, however, that the biological chemistry almost always comes first, and always forms the heart of the discussion – this is by design.

What are the differences between laboratory and biological chemistry?

Although the fundamental principles of organic chemistry are the same regardless of whether we are looking at a biological or laboratory example, there are a few key differences that we will need to keep in mind as we study the subject from both angles. The most important difference between a biological organic reaction and a reaction that takes place in the laboratory is that biological reactions are, almost exclusively, catalyzed by enzymes. The role of an enzyme is both to speed up a reaction and also to guide it, making sure that a specific product results. Laboratory reactions generally take place free in solution (or perhaps on the surface of a metal) without guidance by an enzyme. This means that often more than one possible reaction can take place, and more than one product can form. When we study laboratory reactions, we often have to consider several possible reaction outcomes, learn to predict which one will predominate, and sometimes suggest how a chemist might be able to control reaction conditions so as to maximize the formation of the desired product. When we study enzyme-catalyzed biological reactions, in contrast, one of the central things that we try to understand is how the enzyme is able to guide the reaction so that only the desired product is formed.

Another point to keep in mind is that biological reactions take place exclusively in aqueous (water-based) solution, and at physiological pH and temperature, which for most organisms is pH 7 and about 37^{o} C. In the laboratory, a chemist can run a reaction in many different kinds of solvent, use strong bases or acids, and raise or lower the reaction temperature. One of the most fascinating aspects of studying enzyme-catalyzed organic chemistry is that we can apply our knowledge of the central principles of the subject in order to understand how an enzyme is able to overcome the natural limitations on reaction conditions.

How will this class, with its focus on biological organic chemistry, be different from a class in biochemistry?

You should not make the mistake of thinking that a course built around biological examples is going to be more biochemistry than organic chemistry. Although there will inevitably be some overlap with the biochemistry course you may take later in your career (just as there is overlap between courses in general chemistry and organic chemistry, between physical chemistry and physics, etc.), this is very much an organic chemistry text – it follows essentially the same outline and covers essentially the same topics as most other organic chemistry texts, and is fundamentally distinct from most biochemistry texts. In organic chemistry, we think about individual chemical reactions at the atomic level – concentrating on the 'how and why' details of what is happening with the organic molecule(s) in question. In a biochemistry course, you will see many of the same compounds and reactions that you saw in this text. In biochemistry, however, you will see them mainly from the perspective of complete biochemical pathways: you will spend much of your time studying how these pathways are interconnected and regulated. Biochemistry, then, is a 'bigger picture' course where we study the forest, while in organic chemistry we study the trees.

Chapter 1

Introduction to organic structure and bonding, part I

Introduction

As we begin our study of organic chemistry, there will be a lot of review of things that you have probably learned already from your college or high school General Chemistry course, but there will likely be several concepts that are new to you as well. We will start off with the most basic concept in chemistry: the atom. We'll review some essential facts about the composition of an atom, and then remind ourselves about how electrons occupy 'orbitals' surrounding the nucleus. Then we will start talking about how atoms - particularly atoms of carbon, hydrogen, nitrogen, and oxygen, the most important elements in organic chemistry - share pairs of electrons to form covalent bonds. At first, this discussion will be conducted at the simplest possible level, as we review how to draw 'Lewis structures', assign 'formal charges', and recognize common organic bonding patterns. Later, we will bring in the 'valence bond' theory and the concept of 'hybrid orbitals' in order to help explain the observed three-dimensional geometry of organic structures.

Section 1: Atomic orbitals and electron configuration

1.1A: The atom

Recall from your General Chemistry class that an atom consists of a nucleus, made up of positively charged protons and uncharged neutrons, surrounded by a ‘cloud’ of negatively charged electrons. The number of protons in the nucleus is called the ‘atomic number’, and defines the elemental identity of the atom: if there is only one proton in the nucleus, for example, the atom is a hydrogen, while if there are six protons the atom is a carbon. The **mass number** of an atom is the total number of protons and neutrons in the nucleus.

The number of neutrons in the atomic nucleus can vary, and atoms of the same element that have a different number of neutrons are called **isotopes**. Different isotopes of the same element are differentiated by an isotope number (the number of protons plus the number of neutrons), listed in superscript to the left of the element abbreviation. There are, for example, three main isotopes of hydrogen: about 99% of hydrogen nuclei in nature are composed of simply a single proton, and are designated by the symbol ^{1}H. About 1% of the hydrogen atoms in nature have a single neutron in addition to the proton, and are commonly called deuterium, designated by the symbol ^{2}H. The third hydrogen isotope is tritium, designated by the symbol ^{3}H. A tritium atom has two neutrons in addition to a single proton. Tritium makes up far less than 1% of all hydrogen atoms, and is radioactive. As will see later in this text, both ^{2}H and ^{3}H are commonly used in the laboratory to allow chemists to distinguish between different hydrogen atoms in organic compounds.

The atomic number of carbon is 6, meaning that all carbon nuclei have six protons. The most common isotope of carbon in nature is ^{12}C, with six neutrons. The most common isotope of oxygen (atomic number 8) is ^{16}O, with eight neutrons. Less common carbon and oxygen isotopes that are often used in the organic and biochemistry lab are ^{13}C, ^{14}C (which is radioactive) and ^{18}O.

Exercise 1.1: What is the mass number of:

a) ^{31}P (the most common isotope of phosphorus)?
b) ^{32}P (a radioactive isotope of phosphorus)?
c) ^{37}Cl ?

The protons and neutrons in an atomic nucleus account for virtually all of the atomic mass – the mass of electrons and other subatomic particles is miniscule compared to that of a proton or a neutron. The *size* of an atom, however, is accounted for almost primarily by the space occupied by the electrons circulating around the nucleus – while the diameter of a carbon atom is approximately 7.7×10^{-11} m (0.77 Å), for example, the diameter of the nucleus is only about 10^{-14} m.

While nuclear chemists and physicists think a lot about what is going on inside the nucleus of an atom, most other chemists, organic chemists and biochemists included, worry mainly about what is happening with the electron cloud outside the nucleus. This is because it is the behavior of the electrons that determines how atoms bind to each other to form molecules.

1.1B: Atomic orbitals

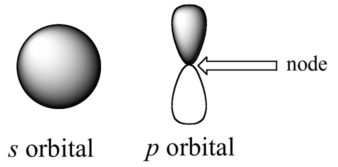
Just how do these electrons behave? You probably remember that chemists talk about electrons occupying ‘**orbitals**’ of different energy levels. Given the use of the term ‘orbital’, it is always tempting for beginning students of chemistry to picture electrons following regular, elliptical orbits around their nucleus, similar to the orbits followed by the planets about the sun, or the moon about the earth. It is very important to remember that this is an incorrect picture – while astronomers can calculate with great precision the path that the moon follows in its orbit, and where it will be located relative to the earth and the sun ten or one hundred years from now, chemists can do no such thing with electrons. With electrons, we can only describe an area in space, called an orbital, where they are *likely* to be.

The branch of science that describes how electrons move around an atomic nucleus is called 'quantum chemistry'. The descriptions and theories that you are about to read may at first sound arbitrary, but rest assured that they are the result of decades of rigorous theoretical research and that they actually do a very good job of describing and explaining the chemical phenomena that scientists observe every day. The details of these theories are far beyond the scope of an introductory organic chemistry course, but you will learn much more if you take a course in physical chemistry or quantum physics.

Electron behavior is described by mathematical equations called **wave equations**, similar to (but much more complicated than) trigonometric equations such as $y = \sin(x)$, that describe regular wave behavior. The solution to an electron's wave equation is called a wavefunction, and is designated by the Greek letter ψ (*psi*). When the value of ψ^2 is plotted in three dimensional space, a volume around the nucleus is defined that is *most likely* to be occupied by the electron. This volume is called an orbital. There is no defined 'edge' to an orbital, but in general chemists draw an orbital as the volume around a nucleus in which there is a 99% chance of finding the electron.

The shape of an orbital is denoted by the letters *s, p, d*, and *f*. In organic chemistry, we are dealing mainly with elements in the second row of the periodic table plus hydrogen. Quantum theory tells us that these elements have only *s* and *p* orbitals.

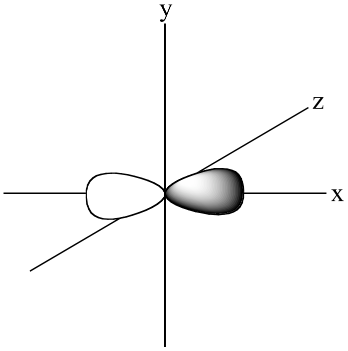
An *s* orbital is a sphere centered on the nucleus. A p orbital, on the other hand, is shaped like a dumb-bell, with two lobes defining where the electrons are likely to be and a '**node'** in the middle that defines a region of zero electron density. The nucleus is located in this node.



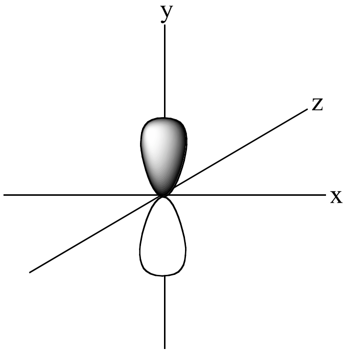
The two lobes of a p orbital have opposite mathematical signs, sometimes indicated with plus and minus signs, and sometimes simply by shading one of the lobes and leaving the other one unshaded (the latter convention will be used throughout this book). Always keep in mind, especially when looking at figures using the plus/minus convention, that the two lobes of a p orbital *do not* have opposite electrical charges like cations and anions (or protons and electrons). Rather, the plus and minus designation simply refers to positive and negative values for the wavefunction, similar to how the solution to the trigonometric function y = sin(x) is sometimes positive and sometimes negative. The opposing signs of the p orbital lobes will have a very real meaning, however, when we later come to a discussion of what is known as 'molecular orbital theory' (section 2.1).

1.1C: Electron configuration

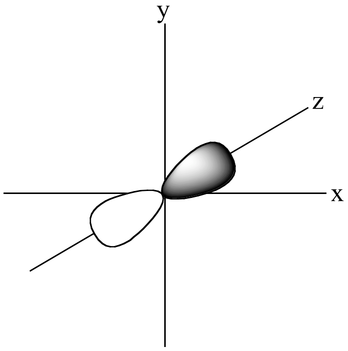
Quantum mechanical calculations tell us that orbitals fall into a series of levels, or shells, of increasing energy and distance from the nucleus. The lowest energy shell, which is also closest to the nucleus, contains only one *s* orbital, designated 1*s*. The second shell, which is higher in energy and further from the nucleus, contains a larger s orbital (2*s*) and three *p* orbitals ($2p_x$, $2p_y$, and $2p_z$). The 2*p* orbitals, which are of equal energy, are oriented perpendicular to one another, along the x, y, and z axes of a three-dimensional coordinate system.



$2p_x$ orbital



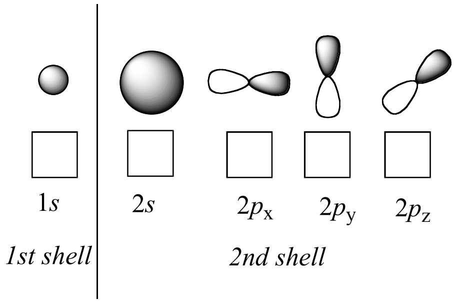
$2p_y$ orbital



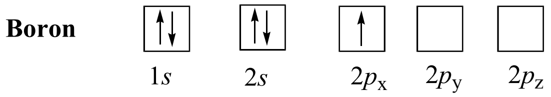
$2p_z$ orbital

Although they are both in the second shell, the 2*s* orbital is lower in energy than the 2*p* orbitals.

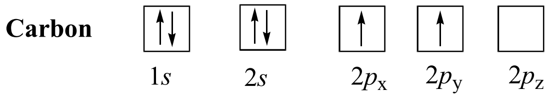
Each orbital can hold two electrons, but they must have opposite 'spins' (recall from general chemistry that each electron in an atom has a quantum 'spin number' of either $m_s = +\frac{1}{2}$ or $-\frac{1}{2}$). When determining the 'electron configuration' for a given atom, you simply need to fill up the orbitals with two electrons each, starting with the lowest energy orbital (1*s*) and continuing on to the higher energy orbitals until all the electrons have been put into an orbital (you may recall that this is called the '*aufbrau*', or 'filling up' principle).



For example, an atom of the element boron, with 5 electrons, has the electron configuration $1s^2\,2s^2\,2p_x^{\,1}$: the 1*s* orbital holds two electrons, the 2*s* holds two more, and the $2p_x$ orbital holds one.



A carbon atom, with six electrons, has the electron configuration $1s^2\,2s^2\,2p_x^{\,1}\,2p_y^{\,1}$.

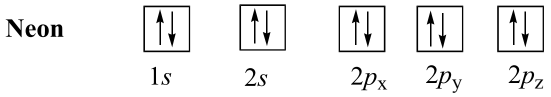


Notice that the sixth electron is put in the $2p_y$ orbital rather than being paired together with the fifth electron in the $2p_x$ orbital. This is the configuration resulting from '**Hund's rule**', which states that the most stable arrangement of electrons in orbitals of equal energy is the one with the greatest number of parallel spins. The $2p_x$ and $2p_y$ electrons in the carbon atom are considered to have parallel spins, whereas if they both occupied the $2p_x$ orbital they would have to have opposite spins.

Exercise 1.2: Illustrate, using the 'box' format as in the figures above, the electron configuration for:

a) a nitrogen atom
b) an oxygen atom
c) a fluorine atom

At the end of the second row of the periodic table, we come to the element neon, which has a 'full shell' of 2*s* and 2*p* orbitals.



In the third row, the *d* orbitals begin to come into play. As stated earlier, we deal in organic chemistry mainly with the second row elements plus hydrogen, so we don't need to consider the *d* orbitals very often. One case where we *will* consider the *d* orbitals is when we discuss, in chapter 10, the covalent bonding situation in the phosphate ion (PO_4^{3-}), and related organic phosphate groups.

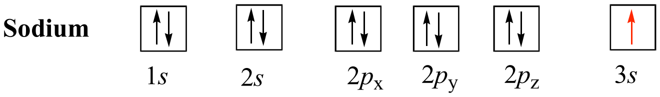
Exercise 1.3: Show the electron configuration for a sulfur atom.

Section 1.2: Chemical Bonds

In our discussion so far we have been talking about isolated atoms – but you are well aware that the study of chemistry mainly deals not with isolated atoms, but with the bonds that form and break between different atoms. Let's review how individual atoms come together to form ionic compounds and covalent molecules and polyatomic ions.

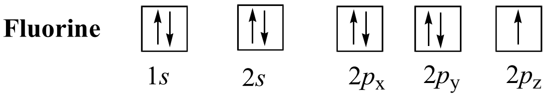
1.2A: Ionic bonds

We'll talk about ionic compounds first. Do you remember the '**octet rule**' from your General Chemistry class? This rule states that the most stable situation for an atom is one in which it has a 'full shell' of electrons, which for our purposes usually means an electron configuration like that of neon, where there are eight electrons filling up the second energy 'shell'. Neon, because it has a full shell of electrons, exists in nature as an isolated atom – in this state, it has already fulfilled the octet rule. Other elements need to lose, gain, or share electrons in order to fulfill the octet rule. Consider the sodium atom:

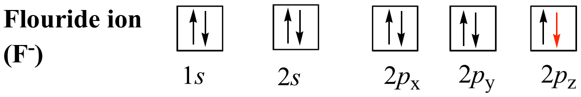


With a full second shell plus one electron in the 3*s* orbital of the third shell, the sodium atom has one electron 'too many' – thus sodium is found in nature mainly as the Na^+ cation, where it has given up its single 3*s* electron to achieve a stable octet configuration that is the same as that of neon. The same can be said of lithium, potassium, and other elements in the first 'alkali metal' row of the periodic table.

The fluorine atom has an electron configuration in which the second shell is one electron shy of holding a full $2s^2\ 2p^6$ octet of electrons:



Thus, fluorine is often found in nature as a chloride anion, F^-, having gained an electron to fulfill the octet rule. Notice that the electron configuration of fluoride ion is the same as that of neon.



The other halogen elements chlorine, bromine, and iodine also are commonly found as anions with a -1 charge.

Exercise 1.4: Give the electron configuration of

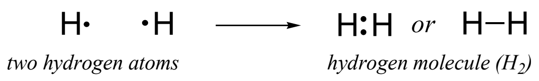
a) a neutral magnesium atom
b) a Mg^{2+} atom
c) a neutral potassium atom
d) a K^+ ion
e) a Cl^- ion

Ionic bonds are formed as the result of electrostatic attraction between two ions with opposite charges. Table salt, for instance, is made up of a crystalline lattice of sodium cations and chloride anions. Although we will primarily be dealing with covalent bonds when we study the structure and reactivity of organic molecules, we will see that electrostatic interactions between oppositely charged species also play very important roles in the reactions of organic compounds.

1.2B: Covalent bonds and Lewis structures

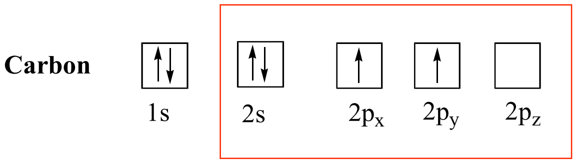
It is now time to discuss the nature of covalent bonds, the predominant bonding type in organic chemistry. We will do this in three stages: first, in this section, we will discuss covalent bonds and covalently bonded molecules and ions in the very simplest of terms, using the 'Lewis structure' convention that you probably recall from General Chemistry. Next, in section 1.5, we will look more closely at the three-dimensional geometry of covalent bonds in organic compounds, making use of the concept of ‘hybrid orbitals’ when discussing ‘valence bond’ theory. Later in section 2.1, we will use another approach, called ‘molecular orbital’ theory, to discuss some phenomena that are not explained well by the first two approaches.

While alkali metals (such as sodium and potassium), alkaline earth metals (such as magnesium and calcium), and halogens (such as fluorine and chlorine) often form ions in order to achieve a full octet, the principle elements of organic chemistry - carbon, hydrogen, nitrogen, and oxygen - instead tend to fill their second shell orbitals by *sharing* electrons with other atoms, forming what we call covalent bonds. Consider the simplest case of hydrogen. An isolated hydrogen atom has only one electron, located in the 1*s* orbital. If two hydrogen atoms come close enough so that their respective 1*s* orbitals overlap, the two electrons can be shared between the two nuclei, and a covalently bonded H_2 molecule is formed.



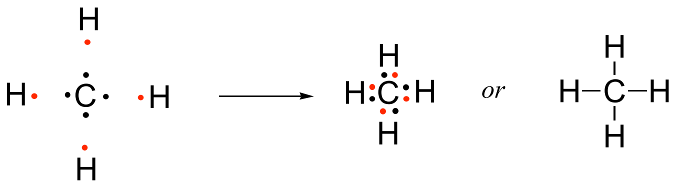
Hydrogen represents is a special case, of course – a hydrogen atom cannot fulfill the octet rule; it needs only two electrons to have a full shell (you could think of this as the 'doublet rule' for hydrogen). In the 'Lewis method', each pair of electrons that is shared between two atoms is drawn as a single line, designating a single covalent bond.

One of the simplest organic molecules is methane, molecular formula CH_4 ((methane is the 'natural gas' burned in home furnaces and hot water heaters, as well as in electrical power generating plants). To illustrate the covalent bonding in methane using the Lewis method, we first must recognize that, although a carbon atom has a total of six electrons, the two electrons in the inner 1*s* orbital do not participate in bonding interactions. It is the other four- those in the 2*s* and 2*p* orbitals - that form covalent bonds with other atoms. Only the partially occupied, highest energy shell of orbitals - in this case the 2*s* and 2*p* orbitals - can overlap with orbitals on other atoms to form covalent bonds. Electrons in these orbitals are termed 'valence electrons'.

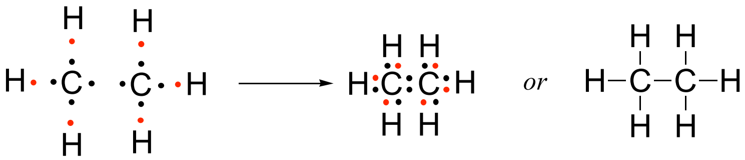


carbon's 4 valence electrons

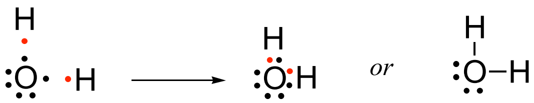
A carbon atom, then, has four valence electrons with which to form covalent bonds. In order to fulfill the octet rule and increase the occupancy of its second shell to eight electrons, it must participate in four electron-sharing interactions - in other words, it must form four covalent bonds. In a methane molecule, the central carbon atom shares its four valence electrons with four hydrogen atoms, thus forming four bonds and fulfilling the octet rule (for the carbon) and the 'doublet rule' (for each of the hydrogens).



The next relatively simple organic molecule to consider is ethane, which has the molecular formula C_2H_6. If we draw each atom with its valence electron(s) separately, we can see that the octet/doublet rule can be fulfilled for all of them by forming one carbon-carbon bond and six carbon-hydrogen bonds.

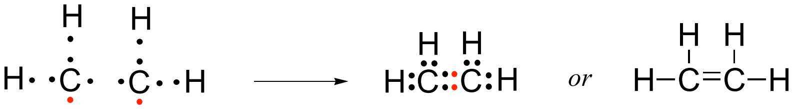


The same approach can be used for molecules in which there is no carbon atom. In a water molecule, two of the six valence electrons on the oxygen atom are used to form bonds to hydrogen atoms, while the remaining four are non-bonding 'lone pairs'.



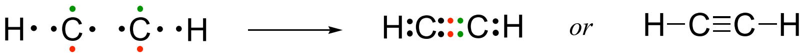
Exercise 1.5: Draw the Lewis structure for ammonia, NH_3.

What about multiple bonds? The molecular formula for ethene (also known as ethylene, a compound found in fruits such as apples that signals them to ripen) is C_2H_4. Arranging the atoms and surrounding them with their valence electrons, you can see that the octet/doublet rule can be fulfilled for all atoms only if the two carbons share *two* pairs of electrons between them - in other words, only if a double bond is formed.

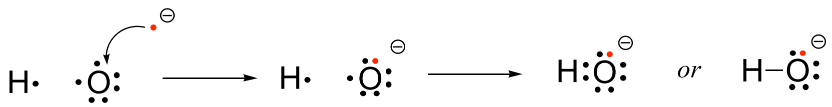


Because a hydrogen atom has only a 1*s* orbital to work with, it cannot form more than one single bond, otherwise it would exceed its doublet rule.

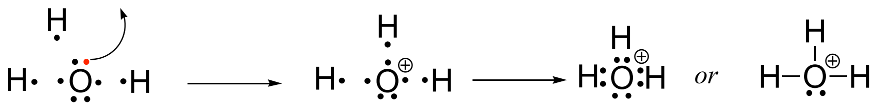
Following this pattern, the triple bond in ethyne molecular formula C_2H_2, (also known as acetylene, the fuel used in welding torches), is formed when the two carbon atoms share *three* pairs of electrons between them.



What about ions? The hydroxide ion, OH^-, is drawn simply by showing the oxygen atom with its six valence electrons, then adding one more electron to account for the negative charge. Now the oxygen has three non-bonding lone pairs, and can only form one bond to a hydrogen. (Bear in mind that this is merely a description of the *thought process* going into drawing a Lewis structure, and is not meant to describe any actual chemical process).



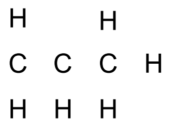
To draw a Lewis structure of the hydronium ion, H_3O^+, you again start with the oxygen atom with its six valence electrons, then take one away to account for the positive charge (there is now one more proton than there are electrons). The oxygen now can form bonds to three hydrogen atoms.



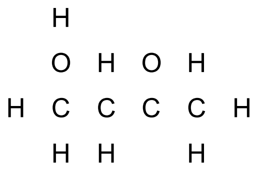
Exercise 1.6: Draw a Lewis structure for the ammonium ion, NH_4^+.

Exercise 1.7: Draw Lewis structures for neutral molecules with the connectivity shown below. Show all lone pairs. The octet rule should be fulfilled for all atoms.

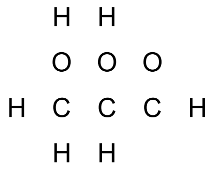
a)



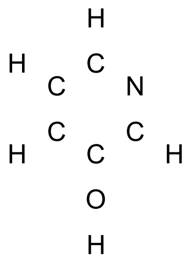
b)



c)



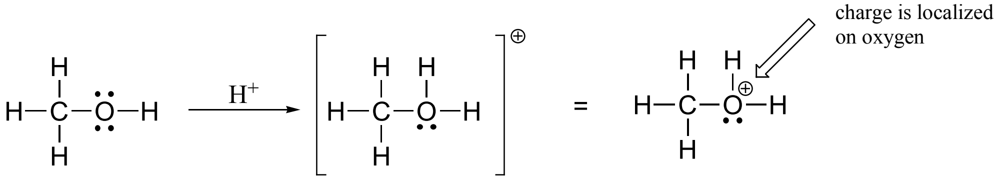
d)



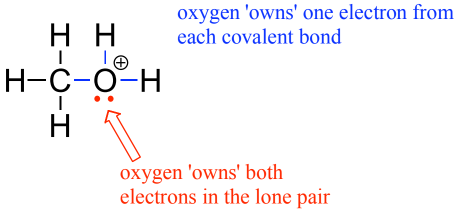
(this molecule contains a six-membered ring)

1.2C: Formal charges

Consider the Lewis structure of methanol, CH_3OH (methanol is the so-called ‘wood alcohol’ that unscrupulous bootleggers sometimes sold during the Prohibition days, often causing the people who drank it to go blind). Just like in a water molecule, the oxygen atom in a methanol molecule has two nonbonding lone pairs of electrons. And just like a water molecule can be protonated to form the H_3O^+ cation, a methanol molecule can be protonated to form $CH_3OH_2^+$ cation.



This polyatomic cation, as you can see, has an overall charge of +1. But we can be more specific than that - we can also state that the positive charge is located specifically on the oxygen atom, rather than on the carbon or any of the hydrogens. When a charge can be located on a particular atom in a polyatomic ion, this atom is said to have a 'formal charge'. Figuring out the formal charge on different atoms of a polyatomic ion is a straightforward process - it's simply a matter of adding up valence electrons. Remember that an oxygen atom needs six valence electrons (in addition to the two electrons in the non-valence 1*s* orbital) to completely balance the charge of the eight protons in its nucleus. Let's figure out how many electrons the oxygen atom in our $CH_3OH_2^+$ ion 'owns'.


oxygen 'owns' one electron from each covalent bond
oxygen 'owns' both electrons in the lone pair

First, we see that there is one lone pair of electrons that the oxygen is not sharing with any other atom - thus it 'owns' both of these electrons. In addition, the oxygen atom is sharing one *pair* of electrons each with three other atoms - since these electrons are shared, we decide that oxygen 'owns' one of each pair, meaning that it owns three bonding electrons. In total then, the oxygen owns five valence electrons: two non-bonding and three bonding. This is one short of the six valence electrons needed to achieve neutrality - thus the oxygen atom has a formal charge of +1.

What is the formal charge, if any, on the other atoms? The carbon needs to 'own' four valence electrons (once again, in addition to the two electrons in its 1*s* orbital) in order to balance the six protons in its nucleus. Because the carbon atom has four single bonds in this structure, and no lone pairs, it does indeed own four electrons (remember that it owns one of each pair of electrons that it shares in a covalent bond). So the carbon has a formal charge of zero.

Each hydrogen needs to own only one electron to balance the charge of its single proton. Indeed, in the $CH_3OH_2^+$ structure, each hydrogen atom has one single bond, meaning that each one owns one electron. Thus all of the hydrogen atoms have formal charges of zero. Notice that, as you would expect, the sum of the formal charges of all the atoms in the ion equals the total charge on the ion - this will always be true for every example you encounter, and is a good way to check to make sure you are figuring individual formal charges correctly.

An abbreviated formula for determining formal charges can be expressed as follows:

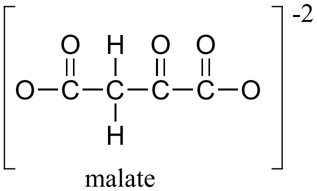
formal charge = (number of valence electrons)
- (number of nonbonding electrons)
- (½ the number of bonding electrons)

When drawing structures, it is very important to show all non-zero formal charges, being clear about where the charges are located. When all non-zero formal charges are shown in the structure, the overall charge on an ion does not need to be indicated - that information is obvious from the sum of the formal charges.

At this point, thinking back to what you learned in general chemistry, you are probably asking "What about dipoles? Doesn't an oxygen atom in an O-H bond 'own' more of the electron density than the hydrogen, because of its greater electronegativity?" You are absolutely correct, and we will be reviewing the concept of dipoles later on. For the purpose of calculating formal charges, however, dipoles don't matter - we always consider the two electrons in a bond to be shared equally, even if that is not an accurate reflection of chemical reality. Formal charges are just that - a formality, a method of electron book-keeping that is tied into the Lewis system for drawing the structures of organic compounds and ions. Later, however, we will see how the concept can help us to visualize how organic molecules react.

Exercise 1.8: Label all non-zero formal charges on the molecules/ions below. All atoms have a full octet of electrons (lone pairs are not shown).

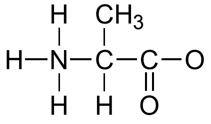
a)



$\left[O{-}C(=O){-}CH_2{-}C(=O){-}C(=O){-}O\right]^{-2}$

malate

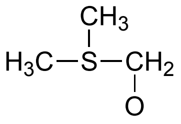
b)



$H{-}NH_2{-}CH(CH_3){-}C(=O){-}O$

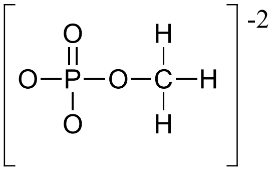
alanine

c)



$H_3C{-}S(CH_3){-}CH_2{-}O$

d)

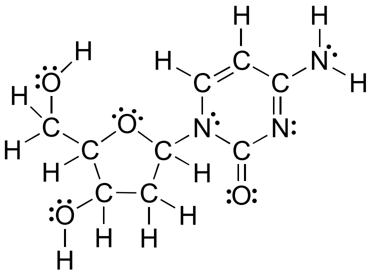


$\left[O{-}P(=O)(O){-}O{-}CH_3\right]^{-2}$

Section 1.3: Drawing organic structures

1.3A: Common bonding patterns in organic structures

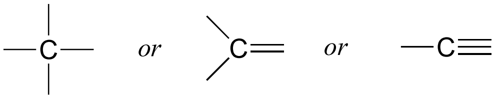
The methods reviewed above for drawing Lewis structures and determining formal charges on atoms are an essential starting point for a novice organic chemist, and work quite will when dealing with small, simple structures. But as you can imagine, these methods become unreasonably tedious and time-consuming when you start dealing with larger structures. It would be unrealistic, for example, to ask you to draw the Lewis structure below (of one of the four nucleoside building blocks that make up DNA) and determine all formal charges by adding up, on an atom-by-atom basis, the valence electrons.



2'-deoxycytidine

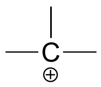
And yet, as organic chemists, and especially as organic chemists dealing with biological molecules, you will be expected soon to draw the structure of large molecules such as this on a regular basis. Clearly, you need to develop the ability to quickly and efficiently draw large structures and determine formal charges. Fortunately, this ability is not terribly hard to come by - all it takes is a few shortcuts and some practice at recognizing common bonding patterns.

Let's start with carbon, the most important element for organic chemists. Carbon is said to be tetravalent, meaning that it tends to form four bonds. If you look at the simple structures of methane, methanol, ethane, ethene, and ethyne in the figures from the previous section, you should quickly recognize that in each molecule, the carbon atom has four bonds, and a formal charge of zero.

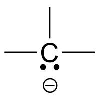


carbon usually has 4 bonds

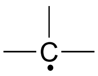
This is a pattern that holds throughout most of the organic molecules we will see, but there are also exceptions. Later on in this chapter and throughout this book we will see examples of organic ions called 'carbocations' and carbanions', in which a carbon atom bears a positive or negative formal charge, respectively. If a carbon has only three bonds and an unfilled valence shell (in other words, if it does not fulfill the octet rule), it will have a positive formal charge.



3 bonds, no lone pair:
carbocation



3 bonds + lone pair:
carbanion



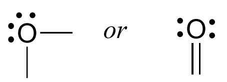
3 bonds + unpaired electron:
carbon radical

If, on the other hand, it has three bonds plus a lone pair of electrons, it will have a formal charge of -1. Another possibility is a carbon with three bonds and a single, unpaired (free radical) electron: in this case, the carbon has a formal charge of zero. (One last possibility is a highly reactive species called a 'carbene', in which a carbon has two bonds and one lone pair of electrons, giving it a formal charge of zero. You may encounter carbenes in more advanced chemistry courses, but they will not be discussed any further in this book).

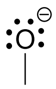
You should certainly use the methods you have learned to check that these formal charges are correct for the examples given above. More importantly, you will need, before you progress much further in your study of organic chemistry, to simply recognize these patterns (and the patterns described below for other atoms) and be able to identify carbons that bear positive and negative formal charges by a quick inspection.

The pattern for hydrogens is easy: hydrogen atoms have only one bond, and no formal charge. The exceptions to this rule are the proton, H^+, and the hydride ion, H^-, which is a proton plus two electrons. Because we are concentrating in this book on organic chemistry as applied to living things, however, we will not be seeing 'naked' protons and hydrides as such, because they are too reactive to be present in that form in aqueous solution. Nonetheless, the *idea* of a proton will be very important when we discuss acid-base chemistry, and the *idea* of a hydride ion will become very important much later in the book when we discuss organic oxidation and reduction reactions. As a rule, though, all hydrogen atoms in organic molecules have one bond, and no formal charge.

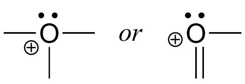
Let us next turn to oxygen atoms. Typically, you will see an oxygen bonding in three ways, all of which fulfill the octet rule.



neutral oxygen: 2 bonds
+ 2 lone pairs



negative formal charge:
1 bond + 3 lone pairs

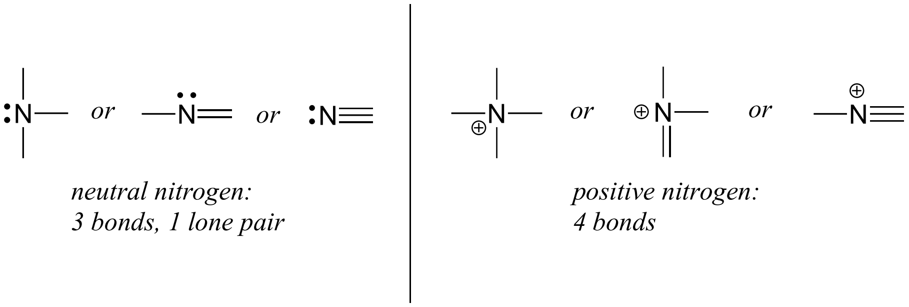


positive formal charge:
3 bonds + 1 lone pair

If it has two bonds and two lone pairs, as in water, it will have a formal charge of zero. If it has one bond and three lone pairs, as in hydroxide ion, it will have a formal charge of -1. If it has three bonds and one lone pair, as in hydronium ion, it will have a formal charge of +1.

When we get to our discussion of free radical chemistry in chapter 17, we will see other possibilities, such as where an oxygen atom has one bond, one lone pair, and one unpaired (free radical) electron, giving it a formal charge of zero. For now, however, concentrate on the three main non-radical examples, as these will account for virtually everything we see until chapter 17.

Nitrogen has two major bonding patterns, both of which fulfill the octet rule:

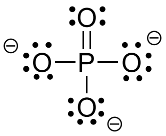


neutral nitrogen:
3 bonds, 1 lone pair

positive nitrogen:
4 bonds

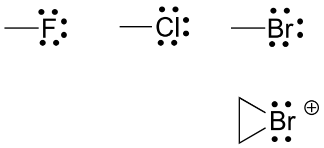
If a nitrogen has three bonds and a lone pair, it has a formal charge of zero. If it has four bonds (and no lone pair), it has a formal charge of +1. In a fairly uncommon bonding pattern, negatively charged nitrogen has two bonds and two lone pairs.

Two third row elements are commonly found in biological organic molecules: sulfur and phosphorus. Although both of these elements have other bonding patterns that are relevant in laboratory chemistry, in a biological context sulfur almost always follows the same bonding/formal charge pattern as oxygen, while phosphorus is present in the form of phosphate ion (PO_4^{3-}), where it has five bonds (almost always to oxygen), no lone pairs, and a formal charge of zero. Remember that elements in the third row of the periodic table have *d* orbitals in their valence shell as well as *s* and *p* orbitals, and thus are not bound by the octet rule.



phosphate

Finally, the halogens (fluorine, chlorine, bromine, and iodine) are very important in laboratory and medicinal organic chemistry, but less common in naturally occurring organic molecules. Halogens in organic compounds usually are seen with one bond, three lone pairs, and a formal charge of zero. Sometimes, especially in the case of bromine, we will encounter reactive species in which the halogen has two bonds (usually in a three-membered ring), two lone pairs, and a formal charge of +1.

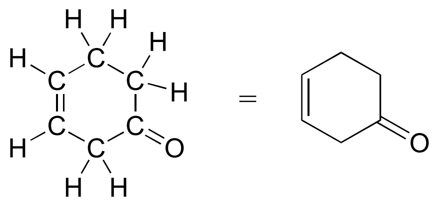


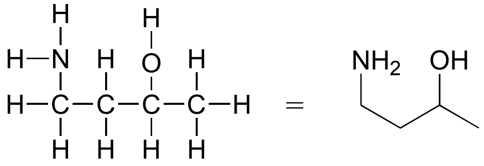
These rules, if learned and internalized so that you don't even need to think about them, will allow you to draw large organic structures, complete with formal charges, quite quickly.

Once you have gotten the hang of drawing Lewis structures, it is not always necessary to draw lone pairs on heteroatoms, as you can assume that the proper number of electrons are present around each atom to match the indicated formal charge (or lack thereof). Occasionally, though, lone pairs are drawn if doing so helps to make an explanation more clear.

1.3B: Using the 'line structure' convention

If you look ahead in this book at the way organic compounds are drawn, you will see that the figures are somewhat different from the Lewis structures we have been using in this section. In order to make structural drawings clearer and less crowded, organic chemists use an abbreviated drawing convention called 'line structures'. The convention is quite simple and makes it easier to draw molecules, but line structures do take a little bit of getting used to. Carbon atoms are depicted not by a capitol C, but by a 'corner' between two bonds, or a free end of a bond. Straight-chain molecules are usually drawn out in a 'zig-zig' shape. Hydrogens attached to carbons are generally not shown: rather, like lone pairs, they are simply implied (unless a positive formal charge is shown, all carbons are assumed to have a full octet of valence electrons). Hydrogens bonded to **heteroatoms** (nitrogen, oxygen, sulfur, or anything other than carbon) *are* shown, but are usually without showing the bond. The following examples illustrate the idea.



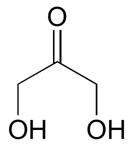


As you can see, the 'pared down' line structure for a molecule makes it much easier to see the basic structure. For larger, more complex biological molecules, it becomes impractical to use full Lewis structures.

Sometimes, one or more carbon atoms in a line structure will be depicted with a capitol C, if doing so makes an explanation easier to follow. If you draw out a carbon with a C, however, you also need to draw in the hydrogens for that carbon.

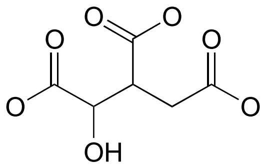
Exercise 1.9: Give the molecular formula, (including overall charge, if any) of the molecules/ions whose structures are shown below, and locate any non-zero formal charges. All atoms (other than hydrogens) have a complete octet of valence electrons.

a)



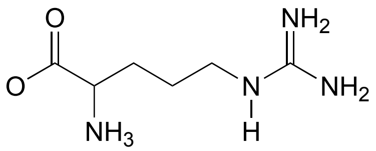
dihydroxyacetone

b)



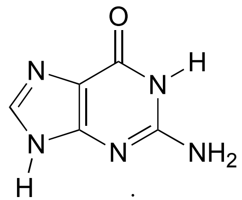
isocitrate

c)



arginine

d)

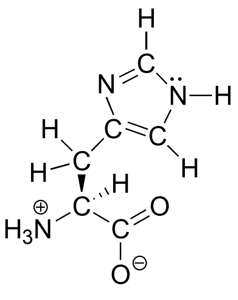


guanine

Exercise 1.10:

a) Draw a line structure for the DNA base 2-deoxycytidine (figure in section 1.3A)

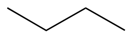
b) Draw a line structure for the amino acid histidine.



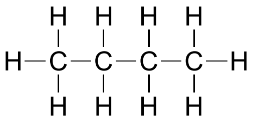
histidine

1.3C: Constitutional isomers

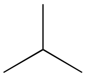
Now that we have reviewed how to draw Lewis structures and learned the line structure shortcut, it is a good time to learn about the concept of constitutional isomers. Imagine if you were asked to draw a structure (Lewis or line) for a compound with the molecular formula C_4H_{10}. This would not be difficult - you could simply draw:



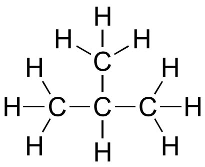
or



But when you compared your answer with that of a classmate, she may have drawn this structure:

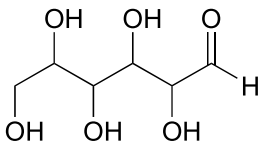


or

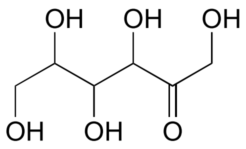


Who is correct? The answer, of course, is that both of you are. A molecular formula only tells you *how many* atoms of each element are present in the compound, not what the actual atom-to-atom connectivity is. There are often many different possible structures for one molecular formula. Compounds that have the same molecular formula but different connectivity are called **constitutional** isomers (sometimes the term '**structural isomer**' is also used). The Greek term '*iso'* means 'same'.

Fructose and glucose, two kinds of sugar molecules, are constitutional isomers with the molecular formula $C_6H_{12}O_6$.



glucose



fructose

Later, we will see other types of isomers that have the same molecular formula and the same connectivity, but are different in other respects.

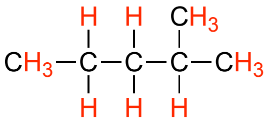
Exercise 1.11 : Draw all of the possible constitutional isomers with the given molecular formula.

a) C_5H_{12}
b) C_4H_{10}
c) C_3H_9N

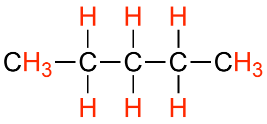
1.3D: The Index of Hydrogen Deficiency

The **Index of Hydrogen Deficiency** (IDH) concept is a helpful tool that we can use when trying to come up with possible Lewis structures when we know a molecular formula. The IHD is really just a simple formula that allows us to calculate how many **units of unsaturation** - in other words, how many multiple bonds or ring structures - exist in a molecule with a given molecular formula.

Here's how it works. To start off with, we will consider only simple hydrocarbons - molecules that contain only carbons and hydrogens. The IHD rule tells us that a saturated, straight-chain alkane will have the molecular formula C_nH_{n+2}, (n is the number of carbons in the molecule). The following alkanes are shown as examples.



C_6H_{14}
$n = 6$
$2n+2 = 14$



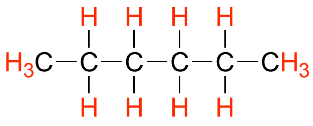
C_5H_{12}
$n = 5$
$2n+2 = 12$

The IHD for hydrocarbons, then, is defined as:

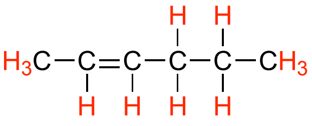
$$\text{IHD} = \frac{(2n+2) - (\#\text{ hydrogens})}{2}$$

By this definition, the IHD value for saturated, straight-chain alkanes is of course equal to zero.

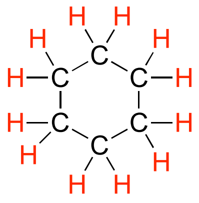
If we put a double bond or a ring structure into a hydrocarbon, the additional carbon-carbon bond means that there must be two fewer hydrogens in the molecule. Because of this, the IHD value is equal to 1.



C_6H_{14}
$2n+2 = 14$
IHD = 0

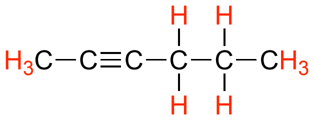


C_6H_{12}
$2n+2 = 14$
IHD = 1



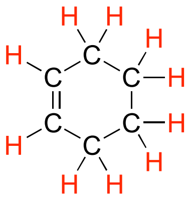
C_6H_{12}
$2n+2 = 14$
IHD = 1

If we put in a double bond and a ring, or a triple bond, then there are two extra carbon-carbon bonds, four fewer hydrogens, and IHD = 2.



$$H_3C{-}C{\equiv}C{-}\overset{H}{\underset{H}{C}}{-}\overset{H}{\underset{H}{C}}{-}CH_3$$

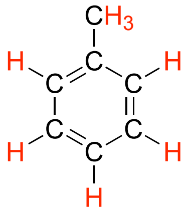
C_6H_{10}
$2n+2 = 14$
IHD = 2



C_6H_{10}
$2n+2 = 14$
IHD = 2

If we are given a molecular formula and asked to come up with one or more possible Lewis structures that fit, figuring the IHD is a very good place to start. For example, if we are asked to draw a possible Lewis structure for a molecule with the molecular formula C_7H_8, we would first calculate that IHD = 4.

$$\boxed{\text{IHD} = \frac{[2(7)+2] - 8}{2} = 4}$$



toluene
(one possible C_7H_8 compound)

We now know that our molecule must have a combination of five multiple bonds or rings. A likely candidate for C_7H_9 would be a compound called toluene (a common solvent in the organic chemistry lab). There are, of course, other constitutional isomers with the same molecular formula.

We can easily modify the IHD formula to account the presence of oxygen, nitrogen, halogens and ionic charges:

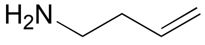
$$\text{IHD} = \frac{(2n+2) - X}{2}$$

where X = (# of hydrogens) + (# of halogens) + (# of negative charges)

- (# of nitrogens) - (# of positive charges)

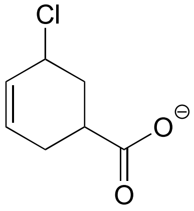
Notice that we may ignore the number of oxygen atoms.

For example, given the molecular formula C_4H_9N, we calculate IHD = 1, so one possible structure is A below. Likewise, given the formula $C_7H_8O_2Cl^-$, we calculate IHD = 3 , for which one possible structure is B below.



A

(one possible C_4H_9N structure)



B

(one possible $C_7H_8O_2Cl$ structure)

Because sulfur and phosphorus are both in the third row of the periodic table and thus able to form more than four bonds, the IHD calculation can get more tricky when these elements are involved. In these cases, we will need to rely on our familiarity with the common organic functional groups (see next section) and use a little more trial and error when suggesting structures. Nevertheless, figuring the IDH will be particularly helpful to us in Chapter 5, when we learn how to use a technique called nuclear magnetic resonance to figure out the structures of unknown organic compounds.

Exercise 1.12 : Determine the IHD for

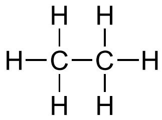
a) 2-deoxycytidine (see section 1.3A for the structure)
b) histidine (exercise 1.10)
c) a molecule with molecular formula $C_9H_{10}NOI$
d) an ion with the molecular formula $C_6H_{11}O_7^-$

Section 1.4: Functional groups and organic nomenclature

1.4A: Common functional groups in organic compounds

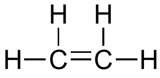
Functional groups are structural units within organic compounds that are defined by specific bonding arrangements between specific atoms. As we progress in our study of organic chemistry, it will become extremely important to be able to quickly recognize the most common functional groups, because they are the key structural elements that define how organic molecules react. For now, we will only worry about drawing and recognizing each functional group, as depicted by Lewis and line structures. Much of the remainder of your study of organic chemistry will be taken up with learning about how the different functional groups tend to behave in organic reactions.

We have already seen some examples of very common functional groups: ethene, for example, contains a carbon-carbon double bond. This double bond is referred to, in the functional group terminology, as an **alkene**.



$H-\underset{H}{\overset{H}{C}}-\underset{H}{\overset{H}{C}}-H$

eth**ane**
(an alk**ane**)



$H-\overset{H}{C}=\overset{H}{C}-H$

eth**ene**
(an alk**ene**)

$H-C\equiv C-H$

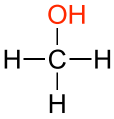
eth**yne**
(an alk**yne**)

The carbon-carbon triple bond in ethyne is the simplest example of an **alkyne** function group.

What about ethane? All we see in this molecule is carbon-hydrogen and carbon-carbon single bonds, so in a sense we can think of ethane as lacking a functional group entirely. However, we do have a general name for this 'default' carbon bonding pattern: molecules or parts of molecules containing only carbon-hydrogen and carbon-carbon single bonds are referred to as **alkanes**.

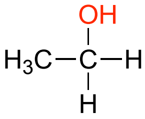
If the carbon of an alkane is bonded to a halogen, the group is now referred to as a **haloalkane** (fluoroalkane, chloroalkane, etc.). Chloroform, $CHCl_3$, is an example of a simple haloalkane.

We have already seen the simplest possible example of an **alcohol** functional group in methanol. In the alcohol functional group, a carbon is single-bonded to an OH group (this OH group, by itself, is referred to as a **hydroxyl**). If the central carbon in an alcohol is bonded to only one other carbon, we call the group a primary alcohol. In secondary alcohols and tertiary alcohols, the central carbon is bonded to two and three carbons, respectively. Methanol, of course, is in class by itself in this respect.



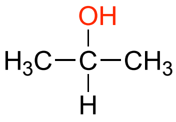
$H-\underset{H}{\overset{OH}{C}}-H$

methanol



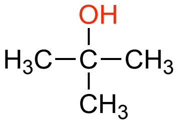
$H_3C-\underset{H}{\overset{OH}{C}}-H$

a primary alcohol



$H_3C-\underset{H}{\overset{OH}{C}}-CH_3$

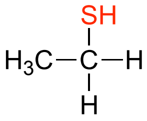
a secondary alcohol



$H_3C-\underset{CH_3}{\overset{OH}{C}}-CH_3$

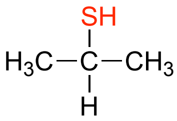
a tertiary alcohol

The sulfur analog of an alcohol is called a **thiol** (the prefix *thio*, derived from the Greek, refers to sulfur).



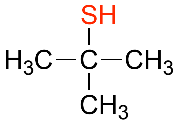
$H_3C-\underset{H}{\overset{SH}{C}}-H$

a primary thiol



$H_3C-\underset{H}{\overset{SH}{C}}-CH_3$

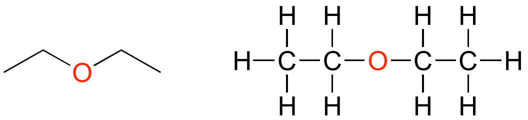
a secondary thiol



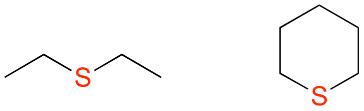
$H_3C-\underset{CH_3}{\overset{SH}{C}}-CH_3$

a tertiary thiol

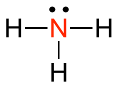
In an **ether** functional group, a central oxygen is bonded to two carbons. Below are the line and Lewis structures of diethyl ether, a common laboratory solvent and also one of the first medical anaesthesia agents.



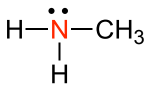
In **sulfides**, the oxygen atom of an ether has been replaced by a sulfur atom.



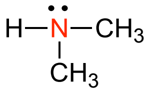
Ammonia is the simplest example of a functional group called **amines**. Just as there are primary, secondary, and tertiary alcohols, there are primary, secondary, and tertiary amines.



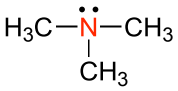
ammonia



a primary amine

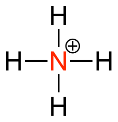


a secondary amine

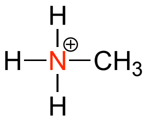


a tertiary amine

One of the most important properties of amines is that they are basic, and are readily protonated to form **ammonium** cations.

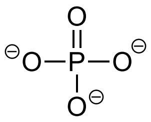


ammonium

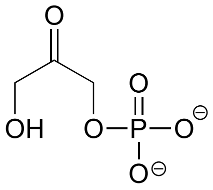


a primary ammonium

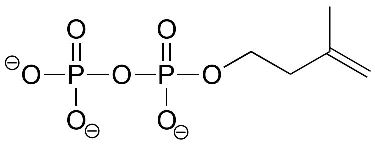
Phosphorus is a very important element in biological organic chemistry, and is found as the central atom in the **phosphate** group. Many biological organic molecules contain phosphate, diphosphate, and triphosphate groups, which are linked to a carbon atom by the **phosphate ester** functionality.



inorganic phosphate

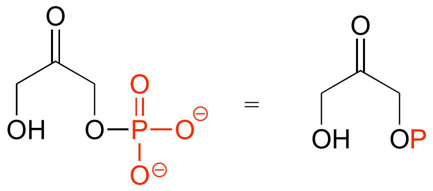


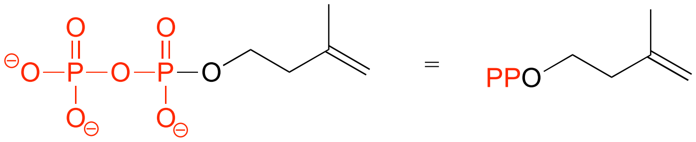
an organic phosphate ester



an organic diphosphate ester

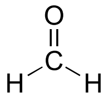
Because phosphates are so abundant in biological organic chemistry, it is convenient to depict them with the abbreviation 'P'. Notice that this 'P' abbreviation includes the oxygen atoms and negative charges associated with the phosphate groups.



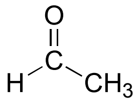


We will have much more to say about the structure of the phosphate group in chapter 10.

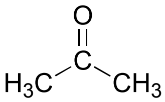
There are a number of functional groups that contain a carbon-oxygen double bond, which is commonly referred to as a **carbonyl**. **Ketones** and **aldehydes** are two closely related carbonyl-based functional groups that react in very similar ways. In a ketone, the carbon atom of a carbonyl is bonded to two other carbons. In an aldehyde, the carbonyl carbon is bonded on one side to a hydrogen, and on the other side to a carbon. The exception to this definition is formaldehyde, in which the carbonyl carbon has bonds to two hydrogens.



formaldehyde

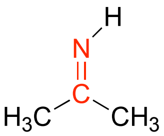


acetaldehyde



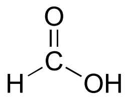
acetone
(a ketone)

Molecules with carbon-nitrogen double bonds are called **imines**, or **Schiff bases**.

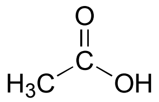


an imine

If a carbonyl carbon is bonded on one side to a carbon (or hydrogen) and on the other side to a **heteroatom** (in organic chemistry, this term generally refers to oxygen, nitrogen, sulfur, or one of the halogens), the functional group is considered to be one of the '**carboxylic acid derivatives'**, a designation that describes a grouping of several functional groups. The eponymous member of this grouping is the **carboxylic acid** functional group, in which the carbonyl is bonded to a hydroxyl (OH) group.

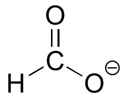


formic acid

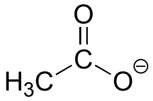


acetic acid
(vinegar)

As the name implies, carboxylic acids are acidic, meaning that they are readily deprotonated to form the conjugate base form, called a **carboxylate** (much more about carboxylic acids in the acid-base chapter!).

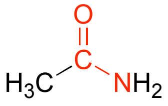


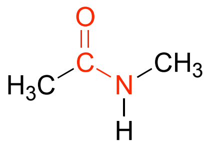
formate

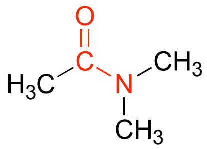


acetate

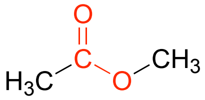
In **amides**, the carbonyl carbon is bonded to a nitrogen. The nitrogen in an amide can be bonded either to hydrogens, to carbons, or to both. Another way of thinking of an amide is that it is a carbonyl bonded to an amine.



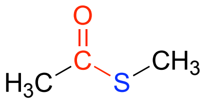




In **esters**, the carbonyl carbon is bonded to an oxygen which is itself bonded to another carbon. Another way of thinking of an ester is that it is a carbonyl bonded to an alcohol. **Thioesters** are similar to esters, except a sulfur is in place of the oxygen.

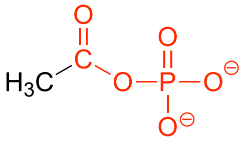


an ester

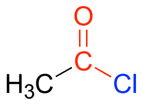


a thioester

In an **acyl phosphate**, the carbonyl carbon is bonded to the oxygen of a phosphate, and in an **acid chloride**, the carbonyl carbon is bonded to a chlorine.



an acyl phosphate



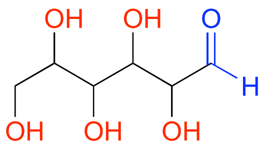
an acid chloride

Finally, in a **nitrile** group, a carbon is triple-bonded to a nitrogen. Nitriles are also often referred to as **cyano** groups.

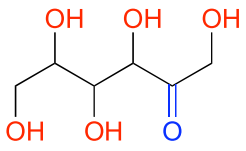
$H_3C-C\equiv N$

a nitrile

A single compound often contains several functional groups. The six-carbon sugar molecules glucose and fructose, for example, contain aldehyde and ketone groups, respectively, and both contain five alcohol groups (a compound with several alcohol groups is often referred to as a ‘**polyol**’).

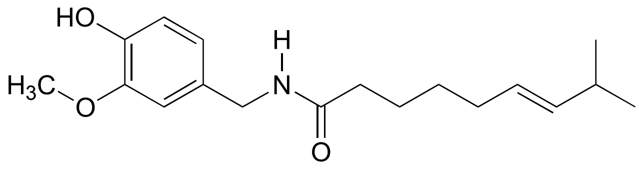


glucose



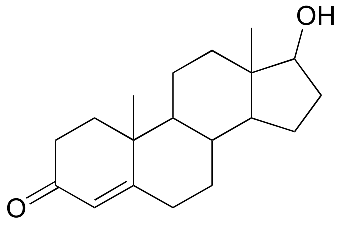
fructose

Capsaicin, the compound responsible for the heat in hot peppers, contains phenol, ether, amide, and alkene functional groups.

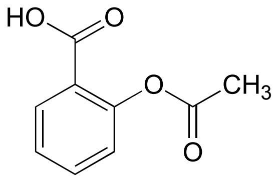


capsiacin

The male sex hormone testosterone contains ketone, alkene, and secondary alcohol groups, while acetylsalicylic acid (aspirin) contains aromatic, carboxylic acid, and ester groups.



testosterone

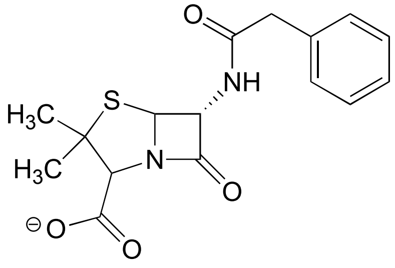


acetylsalicylic acid
(aspirin)

While not in any way a complete list, this section has covered most of the important functional groups that we will encounter in biological and laboratory organic chemistry. The table on the inside back cover provides a summary of all of the groups listed in this section, plus a few more that will be introduced later in the text.

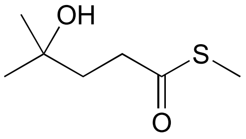
Exercise 1.13: Identify the functional groups in the following organic compounds. State whether alcohols and amines are primary, secondary, or tertiary.

a)

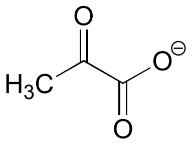


penicillin

b)

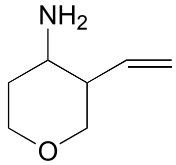


c)



pyruvate

d)



Exercise 1.14: Draw one example each (there are many possible correct answers) of compounds fitting the descriptions below, using line structures. Be sure to designate the location of all non-zero formal charges. All atoms should have complete octets (phosphorus may exceed the octet rule).

a) a compound with molecular formula $C_6H_{11}NO$ that includes alkene, secondary amine, and primary alcohol functional groups

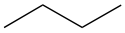
b) an ion with molecular formula $C_3H_5O_6P^{2-}$ that includes aldehyde, secondary alcohol, and phosphate functional groups.

c) A compound with molecular formula C_6H_9NO that has an amide functional group, and does *not* have an alkene group.

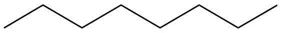
1.4B: Naming organic compounds

A system has been devised by the International Union of Pure and Applied Chemistry (IUPAC, usually pronounced **eye**-you-pack) for naming organic compounds. While the IUPAC system is convenient for naming relatively small, simple organic compounds, it is not terribly useful for the larger, more complex molecules involved in biological organic chemistry, and thus is not generally used in this context. It is, however, a good idea to become familiar with the basic structure of the IUPAC system.

We'll start with the simplest alkane structures. As we already know, CH_4 is called methane, and C_2H_6 ethane. Moving to larger molecule one carbon at a time, we have propane (3 carbons), butane (4 carbons), pentane (5 carbons) and hexane (6 carbons), followed by heptane (7), octane (8), nonane (9), and decane (10).

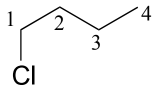


butane



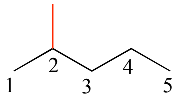
octane

Substituents branching from the main 'parent' chain are located by a carbon number, with the lowest possible number being used.

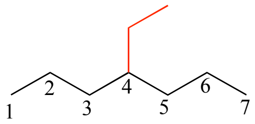


1-chlorobutane

When they are substituents branching from a parent chain, we refer to small alkyl groups with the terms methyl, ethyl, and propyl.

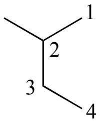


2-methylpentane

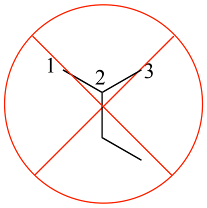


4-ethylheptane

Notice in the example below, the 'ethyl group' is not treated as a substituent, rather it is included as part of the parent chain, and instead the methyl group is treated as a substituent. *The IUPAC name for straight-chain hydrocarbons is always based in the longest possible parent chain.*

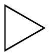


2-methylbutane

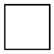


NOT 2-ethylpropane

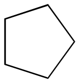
Cyclic structures are called cyclopropane, cyclobutane, cyclopentane, cyclohexane, and so on.



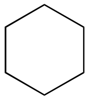
cyclopropane



cyclobutane

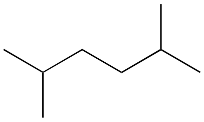


cyclopentane

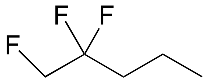


cyclohexane

In the case of multiple substituents, the prefixes *di*, *tri*, and *tetra* are used.

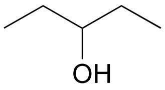


2,5-dimethylhexane

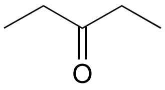


1,2,2,-trifluoropentane

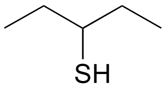
Functional groups have characteristic suffixes. Alcohols, for example, have 'ol' appended to the parent chain name, along with a number designating the location of the hydroxyl group. Ketones are designated by 'one', and thiols by 'thiol'.



3-pentanol

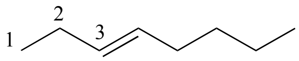


3-pentanone

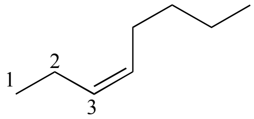


3-pentanethiol

Alkenes are designated with an 'ene' ending. Geometry about the double bond is specified using the terms ***cis*** and ***trans*** (as we will see in section 1.5C, there is no rotation possible about carbon-carbon double bonds, so the two structures below actually represent different molecules). In a *trans* alkene, the alkyl groups are on opposite sides of the double bond, whereas in a *cis* alkene, they are on the same side.



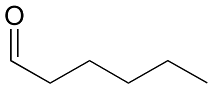
trans-3-octene



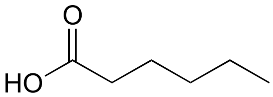
cis-3-octene

We will learn more details about how to specify the arrangement of groups about double bonds in section 3.7C.

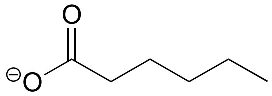
Some groups can only be present on a terminal carbon, and thus a number is not necessary: aldehydes end in 'al', carboxylic acids in 'oic acid', and carboxylates in 'oate'.



hexanal

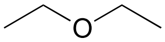


hexanoic acid

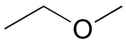


hexanoate

Ethers and sulfides are designated by naming the two groups on either side of the oxygen or sulfur.



diethyl ether

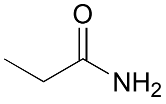


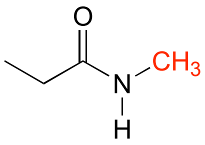
ethyl methyl ether

$H_3C{-}S{-}CH_3$

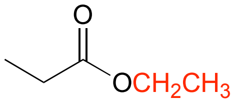
dimethyl sulfide

If an amide has an unsubstituted –NH_2 group, the suffix is simply ‘amide’. In the case of a substituted amide, the group attached to the amide nitrogen is named first, along with the letter ‘N’ to clarify where this group is located.

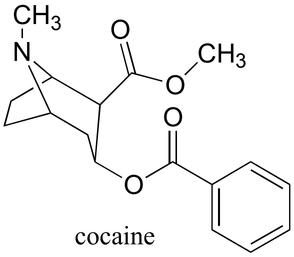

propanamide


N-methylpropanamide

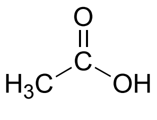
For esters, the group attached to the oxygen is named first, followed by the name of the remaining carboxylate group.

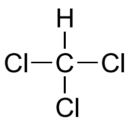

ethyl propanoate

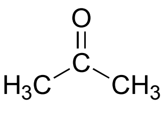
There are of course many more rules in the IUPAC system, and as you can imagine, the IUPAC naming of larger molecules with multiple functional groups and substituents can get very unwieldy very quickly. The illicit drug cocaine, for example, has the IUPAC name 'methyl (1*R*,2*R*,3*S*,5*S*)-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1] octane-2-carboxylate' (this name includes designations for stereochemistry, which is a structural issue that we will not tackle until chapter 3).


cocaine

This is exactly why the IUPAC system is not used for biological organic chemistry - the molecules are just too big. A further complication is that, even outside of a biological context, many simple, common organic molecules are known almost universally by their ‘common’, rather than IUPAC names: acetic acid, chloroform, and acetone are only a few examples.


acetic acid


chloroform


acetone

Exercise 1.15: Give IUPAC names for acetic acid, chloroform, and acetone.

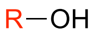
Exercise 1.16: Draw line structures of the following compounds.

a) methylcyclohexane
b) 5-methyl-1-hexanol
c) 2-methyl-2-butene (why is a *cis/trans* designation not necessary?)
d) 5-chloropentanal
e) 2,2-dimethylcyclohexanone
f) *trans*-3-penteneoic acid

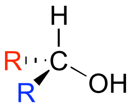
In biochemistry, multiple naming systems are used, depending on whether we are talking about sugars, fats, proteins, or other common biochemical families. Often, nomenclature systems evolve and change as new compound families are discovered. We will not focus very closely in this text on the IUPAC or any other nomenclature system, but if you undertake more advanced study in organic or biological chemistry you may well be expected to learn one or more different systems in some detail.

1.4C: Abbreviated organic structures

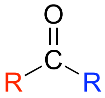
Often when drawing organic structures, chemists find it convenient to use the letter 'R' to designate part of a molecule. If we just want to refer in general to a functional group without drawing a specific molecule, for example, we can use 'R groups' to focus attention on the group of interest:



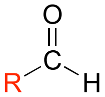
a primary alcohol



a secondary alcohol

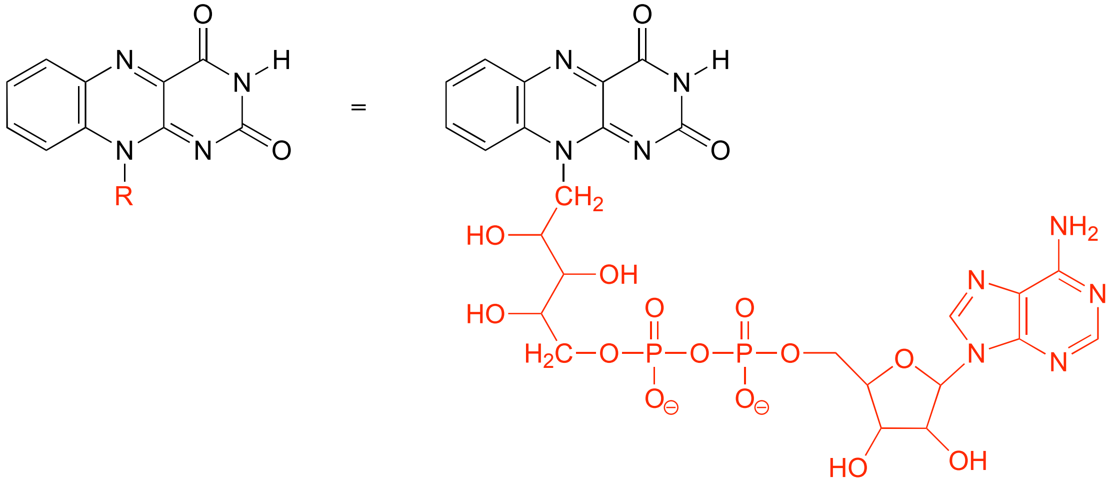


a ketone



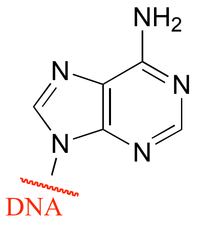
an aldehyde

The R group is also a convenient way to abbreviate the structures of large biological molecules, especially when we are interested in something that is occurring specifically at one location on the molecule.



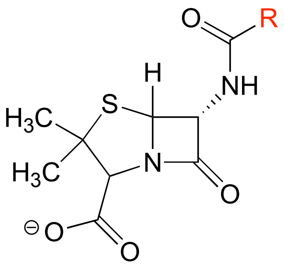
flavin adenine dinucleotide (FAD)

As an alternative, we can use a 'break' symbol to indicate that we are looking at a small piece of a larger molecule. This is used most commonly with in the context of drawing groups on large polymers such as proteins or DNA.

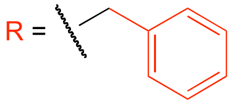


an adenosine in DNA

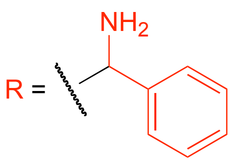
Finally, R groups can be used to concisely illustrate a series of related compounds.



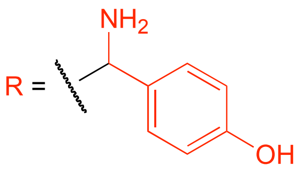
penicillin framework



benzyl penicillin



ampicillin



amoxicillin

Section 1.5: Valence bond theory

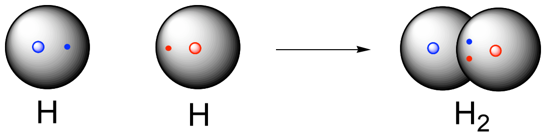
As we have talked about using Lewis structures to depict the bonding in organic compounds, we have been very vague in our language about the actual nature of the chemical bonds themselves. We know that a covalent bond involves the 'sharing' of a pair of electrons between two atoms - but how does this happen, and how does it lead to the formation of a bond holding the two atoms together?

In this section, we will use a model called **valence bond theory** to describe bonding in organic molecules. In this model, bonds are considered to form from the overlapping of two atomic orbitals on different atoms, each orbital containing a single electron. In looking at simple inorganic molecules such as H_2 or HF, our present understanding of s and p atomic orbitals will suffice. In order to explain the bonding in organic molecules, however, we will need to introduce the concept of **hybrid orbitals**.

Valence bond theory is adequate for describing many aspects of organic structure. In some cases, however, chemists need to use a different model, called **molecular orbital (MO) theory**, to talk about covalent bonds in which electrons are shared not just between two atoms, but between several, or even over an entire molcule. The MO model will be introduced in chapter 2 , when these particular chemical phenomena are first discussed. For now, we will concentrate on the more straightforward valence bond theory.

1.5A: Formation of sigma bonds: the H_2 molecule

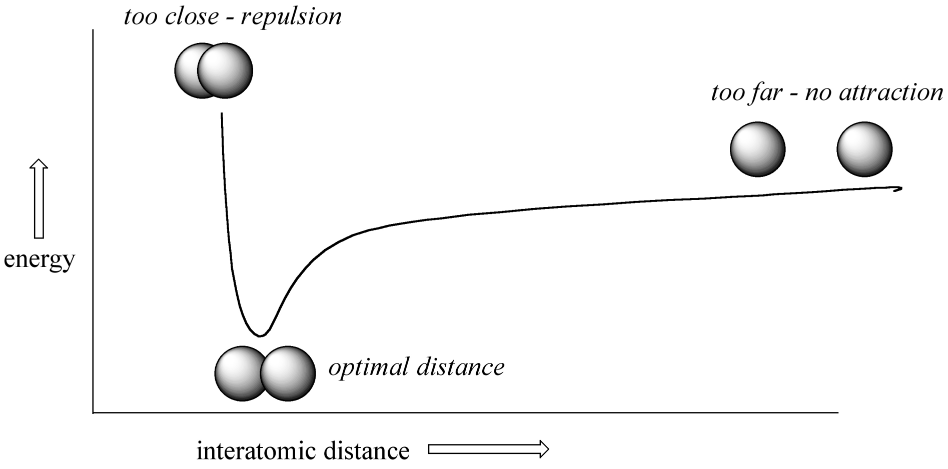
The simplest case to consider is the hydrogen molecule, H_2. When we say that the two electrons from each of the hydrogen atoms are shared to form a covalent bond between the two atoms, what we mean in valence bond theory terms is that the two spherical 1*s* orbitals overlap, allowing the two electrons to form a pair within the two overlapping orbitals.



These two electrons are now attracted to the positive charge of *both* of the hydrogen nuclei, with the result that they serve as a sort of 'chemical glue' holding the two nuclei together.

How far apart are the two nuclei? That is a very important issue to consider. If they are too far apart, their respective 1*s* orbitals cannot overlap, and thus no covalent bond can form - they are still just two separate hydrogen atoms. As they move closer and closer together, orbital overlap begins to occur, and a bond begins to form. This lowers the potential energy of the system, as new, *attractive* positive-negative electrostatic interactions become possible between the nucleus of one atom and the electron of the second.

But something else is happening at the same time: as the atoms get closer, the *repulsive* positive-positive interaction between the two nuclei also begins to increase.



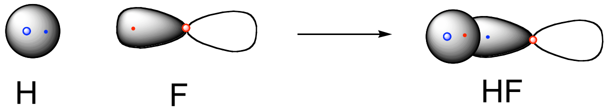
At first this repulsion is more than offset by the attraction between nuclei and electrons, but at a certain point, as the nuclei get even closer, the repulsive forces begin to overcome the attractive forces, and the potential energy of the system rises quickly. When the two nuclei are 'too close', we have a very unstable, high-energy situation. There is a defined optimal distance between the nuclei in which the potential energy is at a minimum, meaning that the combined attractive and repulsive forces add up to the greatest overall attractive force. This optimal internuclear distance is the **bond length**. For the H_2 molecule, this distance is 74×10^{-12} meters, or 0.74 Å (Å means angstrom, or 10^{-10} meters). Likewise, the difference in potential energy between the lowest state (at the optimal internuclear distance) and the state where the two atoms are completely separated is called the **bond energy**. For the hydrogen molecule, this energy is equal to about 104 kcal/mol.

Every covalent bond in a given molecule has a characteristic length and strength. In general, carbon-carbon single bonds are about 1.5 Å long (Å means angstrom, or 10^{-10} meters) while carbon-carbon double bonds are about 1.3 Å, carbon-oxygen double bonds are about 1.2 Å, and carbon-hydrogen bonds are in the range of 1.0 – 1.1 Å. Most covalent bonds in organic molecules range in strength from just under 100 kcal/mole (for a carbon-hydrogen bond in ethane, for example) up to nearly 200 kcal/mole. You can refer to tables in reference books such as the CRC Handbook of Chemistry and Physics for extensive lists of bond lengths, strengths, and many other data for specific organic compounds.

It is not accurate, however, to picture covalent bonds as rigid sticks of unchanging length - rather, it is better to picture them as *springs* which have a defined length when relaxed, but which can be compressed, extended, and bent. This 'springy' picture of covalent bonds will become very important in chapter 4, when we study the analytical technique known as infrared (IR) spectroscopy.

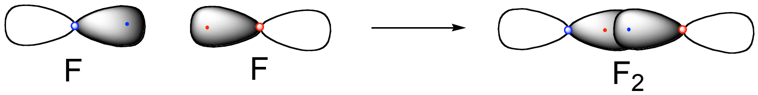
One more characteristic of the covalent bond in H_2 is important to consider at this point. The two overlapping 1*s* orbitals can be visualized as two spherical balloons being squashed together. This means that the bond is *cylindrically symmetrical*: if we were to take a cross-sectional plane of the bond, it would form a circle. This type of bond is referred to as a **σ (sigma) bond**.

We can take a similar approach in using valence orbital theory to describe the covalent bond in hydrofluoric acid, HF. In the electron configuration of an isolated fluorine atom, the 2*s* and two of the three 2*p* orbitals are filled with two electrons each. The third $2p_z$ orbital, however, is occupied only by a single electron, and thus is available for bonding. In the HF molecule, a covalent bond is formed from overlap of the 1*s* orbital in hydrogen with the half-full $2p_z$ orbital in fluorine.



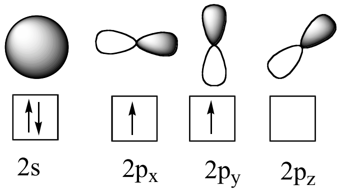
This bond, like that in H_2, has symmetrical symmetry (picture a spherical balloon being pressed directly into one of the ends of an elongated balloon), and thus is classified as a σ bond.

The covalent bond in molecular fluorine, F_2, is treated as the overlap of two half-filled $2p_z$ orbitals, one from each fluorine atom. This also is a σ bond.

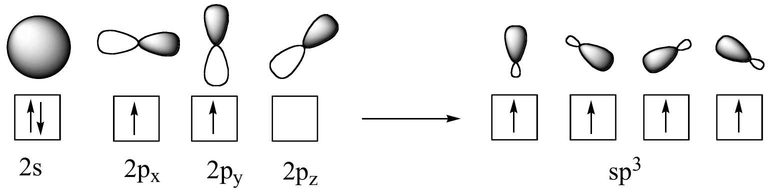


1.5B: Hybrid orbitals - sp^3 hybridization and tetrahedral bonding

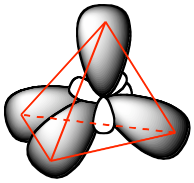
Now let's turn to methane, the simplest organic molecule. Recall the valence electron configuration of the central carbon:



This picture, however, is problematic. How does the carbon form four bonds if it has only two half-filled *p* orbitals available for bonding? A hint comes from the experimental observation that the four C-H bonds in methane are arranged with tetrahedral geometry about the central carbon, and that each bond has the same length and strength. In order to explain this observation, valence bond theory relies on a concept called **orbital hybridization**. In this picture, the four valence orbitals of the carbon (one 2*s* and three 2*p* orbitals) combine mathematically (remember: orbitals are described by equations) to form four equivalent **hybrid orbitals**, which are named **sp^3 orbitals** because they are formed from mixing one s and three p orbitals. In the new electron configuration, each of the four valence electrons on the carbon occupies a single sp^3 orbital.

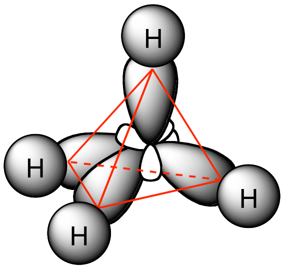


The sp^3 hybrid orbitals, like the *p* orbitals of which they are partially composed, are oblong in shape, and have two lobes of opposite sign. Unlike the *p* orbitals, however, the two lobes are of very different size. The larger lobes of the sp^3 hybrids are directed towards the four corners of a tetrahedron, meaning that the angle between any two orbitals is 109.5°.

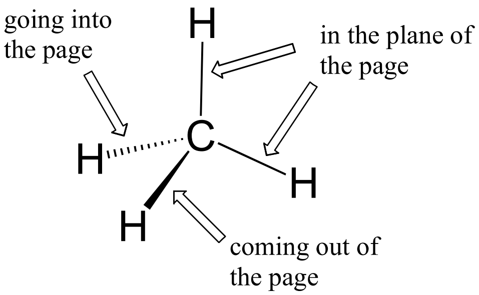


This geometric arrangement makes perfect sense if you consider that it is precisely this angle that allows the four orbitals (and the electrons in them) to be as far apart from each other as possible. This is simply a restatement of the Valence Shell Electron Pair Repulsion (VSEPR) theory that you learned in General Chemistry: electron pairs (in orbitals) will arrange themselves in such a way as to remain as far apart as possible, due to negative-negative electrostatic repulsion.

Each C-H bond in methane, then, can be described as an overlap between a half-filled 1*s* orbital in a hydrogen atom and the larger lobe of one of the four half-filled sp^3 hybrid orbitals in the central carbon. The length of the carbon-hydrogen bonds in methane is 1.09 Å (1.09×10^{-10} m).

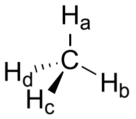


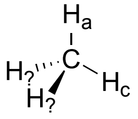
While previously we drew a Lewis structure of methane in two dimensions using lines to denote each covalent bond, we can now draw a more accurate structure in three dimensions, showing the tetrahedral bonding geometry. To do this on a two-dimensional page, though, we need to introduce a new drawing convention: the solid / dashed wedge system. In this convention, a solid wedge simply represents a bond that is meant to be pictured emerging from the plane of the page. A dashed wedge represents a bond that is meant to be pictured pointing into, or behind, the plane of the page. Normal lines imply bonds that lie in the plane of the page.

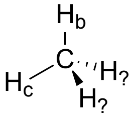


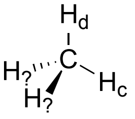
This system takes a little bit of getting used to, but with practice your eye will learn to immediately 'see' the third dimension being depicted.

Exercise 1.17: Imagine that you could distinguish between the four hydrogens in a methane molecule, and labeled them H_a through H_d. In the images below, the *exact same* methane molecule is rotated and flipped in various positions. Draw the missing hydrogen atom labels. (It will be much easier to do this if you make a model.)



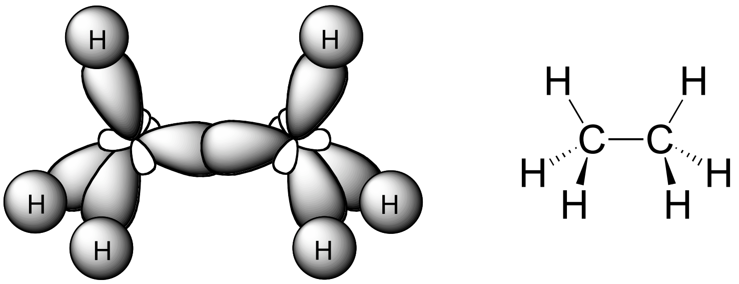
a) 

b) 

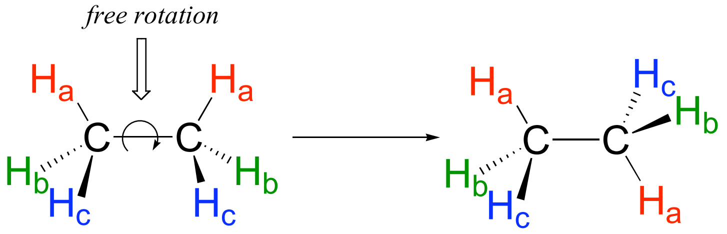
c) 

Exercise 1.18: Describe, with a picture and with words, the bonding in chloroform, $CHCl_3$.

In the ethane molecule, the bonding picture according to valence orbital theory is very similar to that of methane. Both carbons are sp^3-hybridized, meaning that both have four bonds arranged with tetrahedral geometry. The carbon-carbon bond, with a bond length of 1.54 Å, is formed by overlap of one sp^3 orbital from each of the carbons, while the six carbon-hydrogen bonds are formed from overlaps between the remaining sp^3 orbitals on the two carbons and the 1*s* orbitals of hydrogen atoms. All of these are σ bonds.

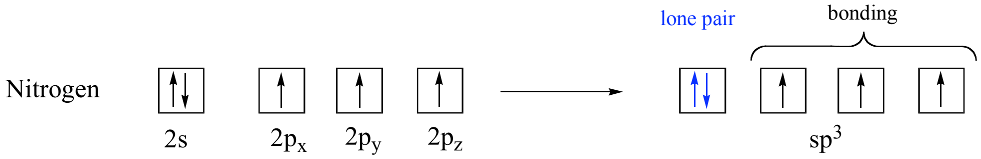


Because they are formed from the end-on-end overlap of two orbitals, *σ bonds are free to rotate*. This means, in the case of ethane molecule, that the two methyl (CH_3) groups can be pictured as two wheels on a hub, each one able to rotate freely with respect to the other.

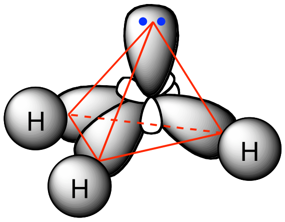


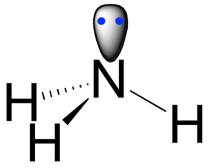
In chapter 3 we will learn more about the implications of rotational freedom in σ bonds, when we discuss the ‘conformation’ of organic molecules.

The sp^3 bonding picture is also used to described the bonding in amines, including ammonia, the simplest amine. Just like the carbon atom in methane, the central nitrogen in ammonia is sp^3-hybridized. With nitrogen, however, there are five rather than four valence electrons to account for, meaning that three of the four hybrid orbitals are half-filled and available for bonding, while the fourth is fully occupied by a (non-bonding) pair of electrons.



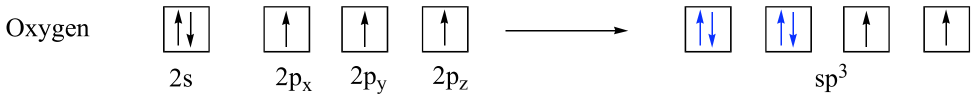
The bonding arrangement here is also tetrahedral: the three N-H bonds of ammonia can be pictured as forming the base of a trigonal pyramid, with the fourth orbital, containing the lone pair, forming the top of the pyramid.

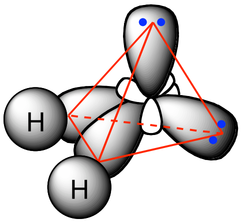


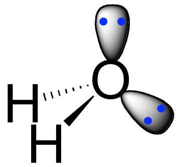


Recall from your study of VSEPR theory in General Chemistry that the lone pair, with its slightly greater repulsive effect, ‘pushes’ the three N-H σ bonds away from the top of the pyramid, meaning that the H-N-H bond angles are slightly less than tetrahedral, at 107.3° rather than 109.5°.

VSEPR theory also predicts, accurately, that a water molecule is ‘bent’ at an angle of approximately 104.5°. It would seem logical, then, to describe the bonding in water as occurring through the overlap of sp^3-hybrid orbitals on oxygen with 1*s* orbitals on the two hydrogen atoms. In this model, the two nonbonding lone pairs on oxygen would be located in sp^3 orbitals.







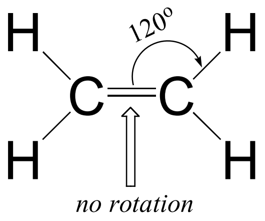
Some experimental evidence, however, suggests that the bonding orbitals on the oxygen are actually unhybridized 2*p* orbitals rather than sp^3 hybrids. Although this would seem to imply that the H-O-H bond angle should be 90° (remember that *p* orbitals are oriented perpendicular to one another), it appears that electrostatic repulsion has the effect of distorting this *p*-orbital angle to 104.5°. Both the hybrid orbital and the nonhybrid orbital models present reasonable explanations for the observed bonding arrangement in water, so we will not concern ourselves any further with the distinction.

Exercise 1.19: Draw, in the same style as the figures above, an orbital picture for the bonding in methylamine.

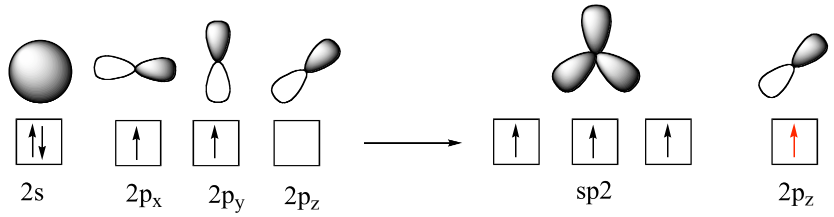
1.5C: Formation of pi bonds - sp^2 and sp hybridization

The valence bond theory, along with the hybrid orbital concept, does a very good job of describing double-bonded compounds such as ethene. Three experimentally observable characteristics of the ethene molecule need to be accounted for by a bonding model:

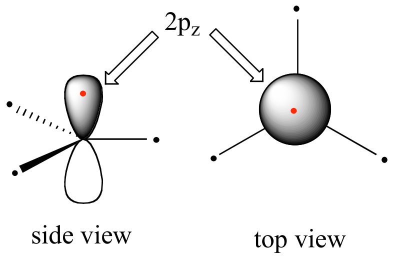
1) Ethene is a planar (flat) molecule.
2) Bond angles in ethene are approximately 120°, and the carbon-carbon bond length is 1.34 Å, significantly shorter than the 1.54 Å single carbon-carbon bond in ethane.
3) There is a significant barrier to rotation about the carbon-carbon double bond.



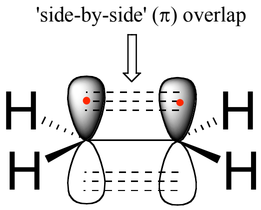
Clearly, these characteristics are not consistent with an sp^3 hybrid bonding picture for the two carbon atoms. Instead, the bonding in ethene is described by a model involving the participation of a different kind of hybrid orbital. Three atomic orbitals on each carbon – the $2s$, $2p_x$ and $2p_y$ orbitals – combine to form three 'sp^2' hybrids, leaving the $2p_z$ orbital unhybridized.



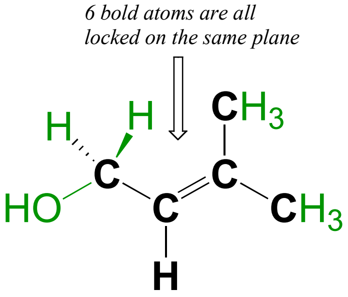
The three sp^2 hybrids are arranged with trigonal planar geometry, pointing to the three corners of an equilateral triangle, with angles of 120° between them. The unhybridized $2p_z$ orbital is *perpendicular* to this plane (in the next several figures, sp^2 orbitals and the σ bonds to which they contribute are represented by lines and wedges; only the $2p_z$ orbitals are shown in the 'space-filling' mode).



The carbon-carbon double bond in ethene consists of one σ bond, formed by the overlap of two sp^2 orbitals, and a second bond, called **a π (pi) bond**, which is formed by the *side-by-side* overlap of the two unhybridized $2p_z$ orbitals from each carbon.

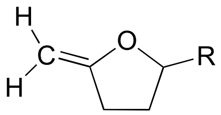


The π bond does *not* have symmetrical symmetry. Because it is the result of side-by-side overlap (rather then end-to-end overlap like a σ bond), rotation about the carbon-carbon double bond is very difficult. If rotation about this bond were to occur, it would involve disrupting the side-by-side overlap between the two $2p_z$ orbitals that make up the π bond. The presence of the π bond thus 'locks' the six atoms of ethene into the same plane. This argument extends to larger alkene groups: in each case, the six atoms of the group form a single plane.

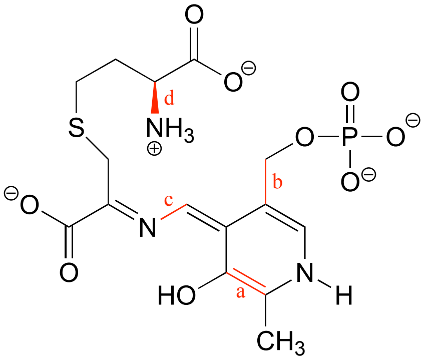


Conversely, σ bonds such as the carbon-carbon single bond in ethane (CH_3CH_3) exhibit free rotation, and can assume many different conformations, or shapes - this is one of the main subjects of Chapter 3.

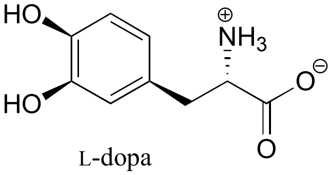
Exercise 1.20: Circle the six atoms in the molecule below that are 'locked' into the same plane.



Exercise 1.21: What kinds of orbitals are overlapping in bonds a-d indicated below?

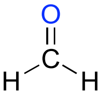


Exercise 1.22: What is wrong with the way the following structure is drawn?

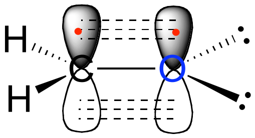


L-dopa

A similar picture can be drawn for the bonding in carbonyl groups, such as formaldehyde. In this molecule, the carbon is sp^2-hybridized, and we will assume that the oxygen atom is also sp^2 hybridized. The carbon has three σ bonds: two are formed by overlap between two of its sp^2 orbitals with the 1*s* orbital from each of the hydrogens, and the third σ bond is formed by overlap between the remaining carbon sp^2 orbital and an sp^2 orbital on the oxygen. The two lone pairs on oxygen occupy its other two sp^2 orbitals.

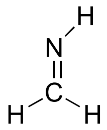


formaldehyde



The π bond is formed by side-by-side overlap of the unhybridized $2p_z$ orbitals on the carbon and the oxygen. Just like in alkenes, the $2p_z$ orbitals that form the π bond are perpendicular to the plane formed by the σ bonds.

Exercise 1.23: Describe and draw the bonding picture for the imine group shown below. Use the drawing of formaldehyde above as your guide.

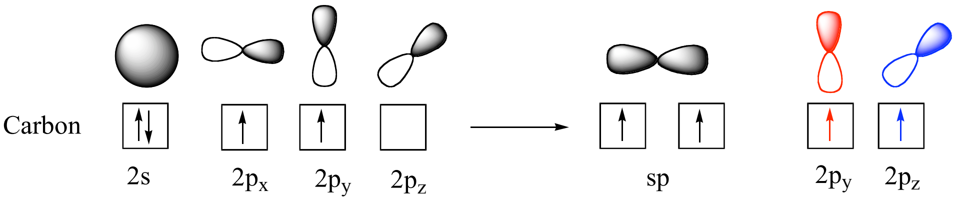


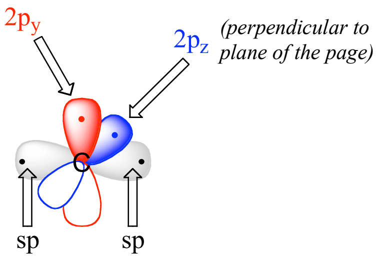
Finally, the hybrid orbital concept applies well to triple-bonded groups, such as alkynes and nitriles. Consider, for example, the structure of ethyne (common name acetylene), the simplest alkyne.

$$H-C\equiv C-H$$

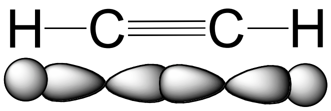
ethyne
(acetylene)

This molecule is linear: all four atoms lie in a straight line. The carbon-carbon triple bond is only 1.20Å long. In the hybrid orbital picture of acetylene, both carbons are **sp-hybridized**. In an sp-hybridized carbon, the 2*s* orbital combines with the $2p_x$ orbital to form two sp hybrid orbitals that are oriented at an angle of 180° with respect to each other (eg. along the x axis). The $2p_y$ and $2p_z$ orbitals remain unhybridized, and are oriented perpendicularly along the y and z axes, respectively.

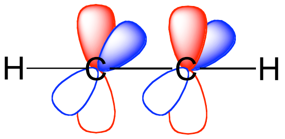




The C-C σ bond, then, is formed by the overlap of one sp orbital from each of the carbons, while the two C-H σ bonds are formed by the overlap of the second sp orbital on each carbon with a 1*s* orbital on a hydrogen. Each carbon atom still has two half-filled $2p_y$ and $2p_z$ orbitals, which are perpendicular both to each other and to the line formed by the σ bonds. These two perpendicular pairs of *p* orbitals form two π bonds between the carbons, resulting in a triple bond overall (one σ bond plus two π bonds).

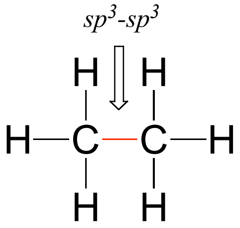


σ bonding framework

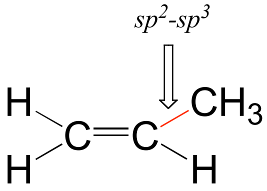


π overlap

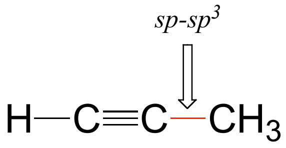
The hybrid orbital concept nicely explains another experimental observation: single bonds adjacent to double and triple bonds are progressively shorter and stronger than 'normal' single bonds, such as the one in a simple alkane. The carbon-carbon bond in ethane (structure A below) results from the overlap of two sp^{3} orbitals.



A



B



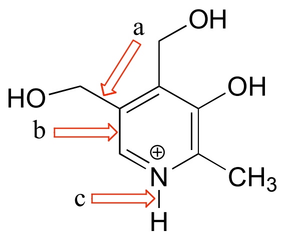
C

In alkene B, however, the carbon-carbon single bond is the result of overlap between an sp^{2} orbital and an sp^{3} orbital, while in alkyne C the carbon-carbon single bond is the result of overlap between an sp orbital and an sp^{3} orbital. These are all σ bonds, but the bond in C is shorter and stronger than the one in B, which is in turn shorter and stronger than the one in A.

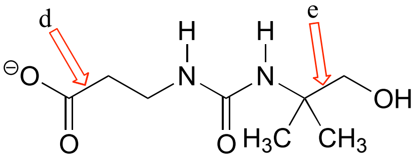
The explanation here is relatively straightforward. An sp orbital is composed of one *s* orbital and one *p* orbital, and thus it has 50% *s* character and 50% *p* character. sp^{2} orbitals, by comparison, have 33% *s* character and 67% *p* character, while sp^{3} orbitals have 25% *s* character and 75% *p* character. Because of their spherical shape, 2*s* orbitals are smaller, and hold electrons closer and 'tighter' to the nucleus, compared to 2*p* orbitals. Consequently, bonds involving sp + sp^{3} overlap (as in alkyne C) are shorter and stronger than bonds involving sp^{2} + sp^{3} overlap (as in alkene B). Bonds involving sp^{3}-sp^{3} overlap (as in alkane A) are the longest and weakest of the group, because of the 75% '*p*' character of the hybrids.

Exercise 1.24: For each of the bonds indicated by arrows a-e in the figures below, describe the bonding picture by completing this sentence:

"The σ bond indicated by this arrow is formed by the overlap of an ________ orbital of a _________atom and an ________orbital of a _______atom."



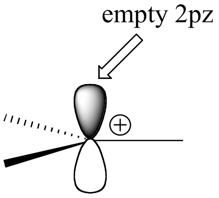
pyridoxine
(vitamin B_6)



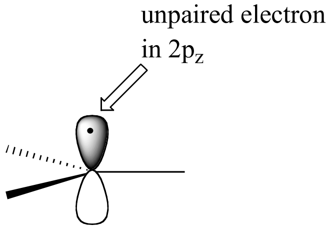
pantothenate
(vitamin B_5)

1.5D: The valence bonding picture in carbocations, carbanions, and carbon free radicals

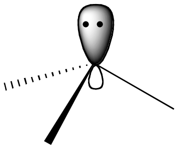
Although they are almost always high-energy, unstable species, carbocations, carbanions, and carbon free radical species play important - if short-lived – roles in many organic reactions. If a carbon atom bears a positive formal charge, or has an unpaired electron, it is generally considered to be sp^2 hybridized, meaning that its three σ bonds are arranged with trigonal planar geometry. That just leaves the perpendicular, unhybridized $2p_z$ orbital to account for. If the carbon bears a positive charge, the $2p_z$ orbital is considered to be empty. In the case of a radical, the $2p_z$ orbital is occupied by a single unpaired electron.



bonding in a carbocation



bonding in a carbon radical



bonding in a methyl carbanion

The picture for a negatively-charged carbon - a carbanion - depends on the broader molecular context, specifically the presence of conjugated π bonds (section 2.1B). In the simple case of a methyl anion, however, however, the bonding is pictured as sp^3, with one of the four orbitals holding the lone pair of electrons.

The orbital and bonding arrangements for carbocation, radicals, and carbanions will be very important to consider when we begin to consider the idea of resonance (section 2.2).

Chapter 1 problems

P1.1: Draw a picture showing the orbitals involved in bonding in the molecules below. Draw *all* bonds, both σ and π, as overlapping orbitals. Locate all lone pairs in their appropriate orbitals.

a) dimethyl ether (CH_3OCH_3)
b) ethanol (CH_3CH_2OH)
c) acetaldehyde (CH_3COH)
d) hydrogen cyanide (HCN)

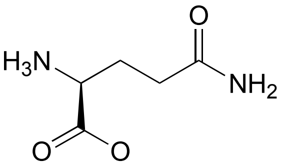
P1.2: Give the electron configuration for:

a) a lithium cation (Li^+)
b) a calcium cation (Ca^{2+})
c) a iron cation in the ferric (Fe^{+3}) state

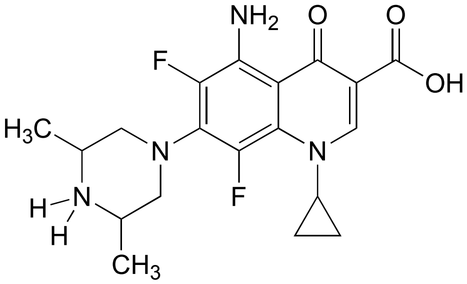
P1.3: Draw one example each (there are many possible correct answers) of compounds fitting the descriptions below, using line structures. Be sure to designate the location of all non-zero formal charges. All atoms should have complete octets (phosphorus may exceed the octet rule).

a) an 8-carbon molecule with secondary alcohol, primary amine, amide, and *cis*-alkene groups
b) a 12-carbon molecule with carboxylate, diphosphate, and lactone (cyclic ester) groups.
c) a 9-carbon molecule with cyclopentane, *trans*-alkene, ether, and aldehyde groups

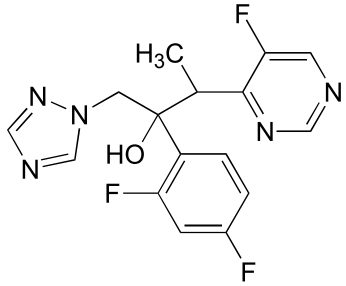
P1.4: For the structures below, label all non-zero formal charges, and determine the molecular formula and the IHD.



glutamine, an amino acid
(formal charges not shown)



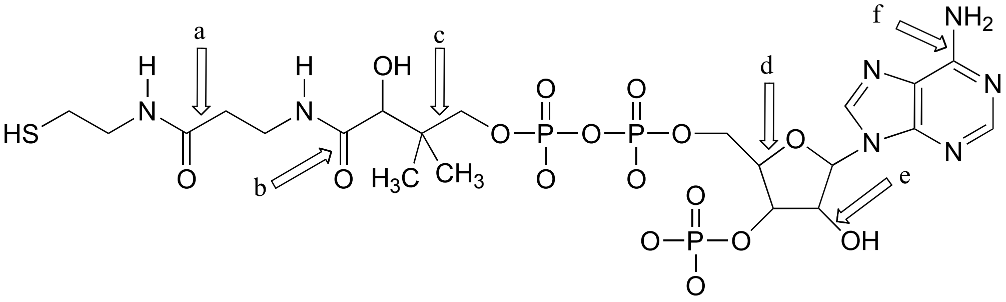
Sparfloxacin, a DNA gyrase inhibitor antibiotic
(formal charges not shown)



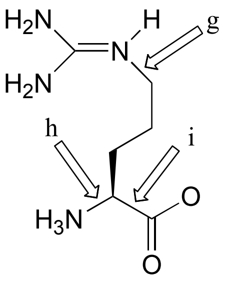
Voriconazol, an antifungal agent

P1.5: In the structures shown below:

a-i) Describe the orbitals involved in the bonds indicated by the arrows.
j) Fill in all formal charges
k) Give the molecular formula for arginine



Coenzyme A
(formal charges not shown)



arginine
(formal charges not shown)

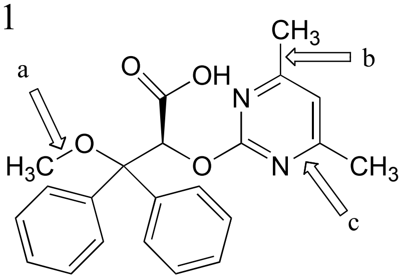
P1.6: The four compounds below appeared in the October 9 and October 25, 2006 issues of Chemical and Engineering News.

a-k) For each bond indicated by an arrow, specify the types of orbitals that are overlapping.

l) Which compound contains two aldehydes? Which contains an ether? Which contains an amide? Which contains a terminal alkene? Which contains an amine (and is this amine primary, secondary, tertiary, or quaternary?)

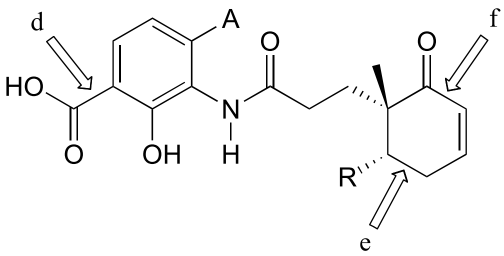
m) Give the molecular formula for compound 3

1



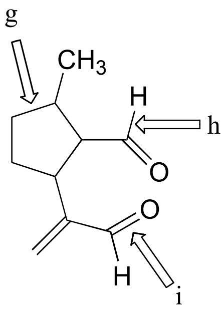
a hypertension drug

2



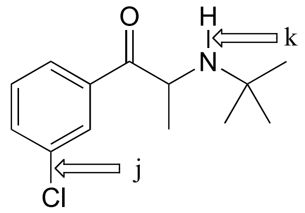
a newly-discovered antibiotic compound

3



a defensive chemical extracted from a walking stick insect

4

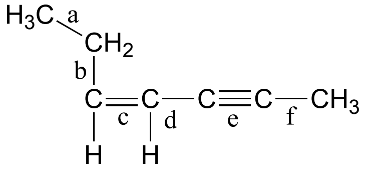


a drug for smoking cessation

P1.7: Draw four isomers with the molecular formula C_4H_8.

P1.8: Draw structures of four different amides with molecular formula C_3H_7NO.

P1.9: Rank the bonds a-f below according to increasing bond length.



P1.10: Draw the structures of the following organic molecules:

a) 2,6-dimethyldecane
b) 2,2,5,5-tetramethyl-3-hexanol
c) methyl butanoate
d) N-ethylhexanamide
e) 7-fluoroheptanoate
f)) 1-ethyl-3,3-dimethylcyclohexene

Challenge problems

C1.1: Imagine that you hear a description of the bonding in water as being derived directly from the atomic orbital theory, without use of the hybrid orbital concept. In other words, the two bonds would be formed by the overlap of the half-filled $2p_y$ and $2p_z$ orbitals of oxygen with the 1s orbitals of hydrogen, while the two lone pairs on oxygen would occupy the 2s and $2p_x$ orbitals. What is wrong with this picture? How would the bonding geometry differ from what is actually observed for water?

C1.2: Draw a picture showing the geometry of the overlapping orbitals that form the bonding network in allene, H_2CCCH_2. Then, draw a Lewis structure for the molecule, using the solid/dash wedge bond convention as necessary to indicate the correct geometry of the σ bonds.

Chapter 2

Introduction to organic structure and bonding, part II

Introduction

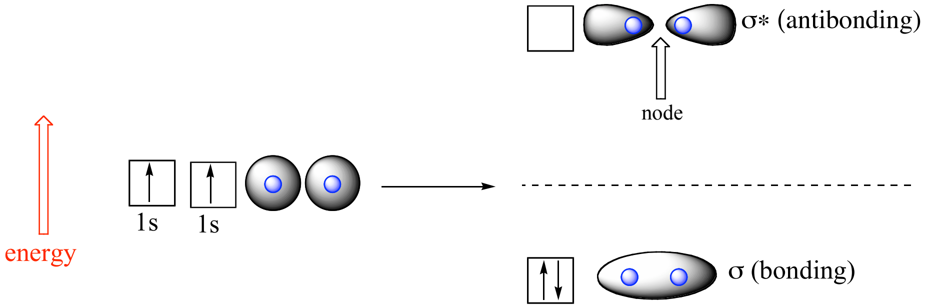
In this chapter, we will continue our discussion of organic structure. Just like in chapter 1, many of the concepts we talk about here will be familiar to you from General Chemistry, but we will also venture into new territory on plenty of occasions. To begin with, we will review the 'molecular orbital' (MO) theory of bonding, and see how this way of thinking about chemical structure can help us to understand the behavior of molecules with one or more double bonds. As part of our exploration of MO theory, we will look at the unique properties of 'aromatic' groups, and learn how to use the 'resonance contributor' convention to depict situations where electrons are spread out over several atoms in a molecule rather than being localized on one atom. Later, we will shift our focus to noncovalent interactions, and review how these forces are responsible for many of the physical properties of compounds that we can observe and measure.

Section 2.1: Molecular orbital theory

As we have seen, valence bond theory does a remarkably good job at explaining the bonding geometry of many of the functional groups in organic compounds. There are some areas, however, where the valence bond theory falls short. It fails to adequately account, for example, for some interesting properties of compounds that contain alternating double and single bonds. In order to understand these properties, we need to think about chemical bonding in a new way, using the ideas of molecular orbital (MO) theory.

2.1A: Another look at the H_2 molecule: bonding and antibonding sigma molecular orbitals

Let's go back and consider again the simplest possible covalent bond: the one in molecular hydrogen (H_2). When we described the hydrogen molecule using valence bond theory, we said that the two 1*s* orbitals from each atom overlap, allowing the two electrons to be shared and thus forming a covalent bond. In molecular orbital theory, we make a further statement: we say that the two atomic 1*s* orbitals don't just overlap, they actually *combine to form two completely new orbitals*. These two new orbitals, instead of describing the likely location of an electron around a single nucleus, describe the location of an electron pair around two or more nuclei. The bonding in H_2, then, is due to the formation of a new **molecular orbital** (MO), specifically a σ molecular orbital, in which a pair of electrons is delocalized around two hydrogen nuclei. An important principle of quantum mechanical theory is that when orbitals combine, the number of orbitals before the combination takes place must equal the number of new orbitals that result – orbitals don't just disappear! (we saw this previously when we discussed hybrid orbitals: one *s* and three *p* orbitals make four sp^3 hybrids). When two atomic 1*s* orbitals combine in the formation of H_2, the result is two molecular orbitals called **sigma (σ) orbitals**. According to MO theory, the first σ orbital is lower in energy than either of the two isolated atomic 1*s* orbitals – thus this σ orbital is referred to as a **bonding molecular orbital**. The second, **σ* orbital** is higher in energy than the two atomic 1*s* orbitals, and is referred to as an **antibonding molecular orbital** (in MO theory, a star (*) sign always indicates an antibonding orbital).



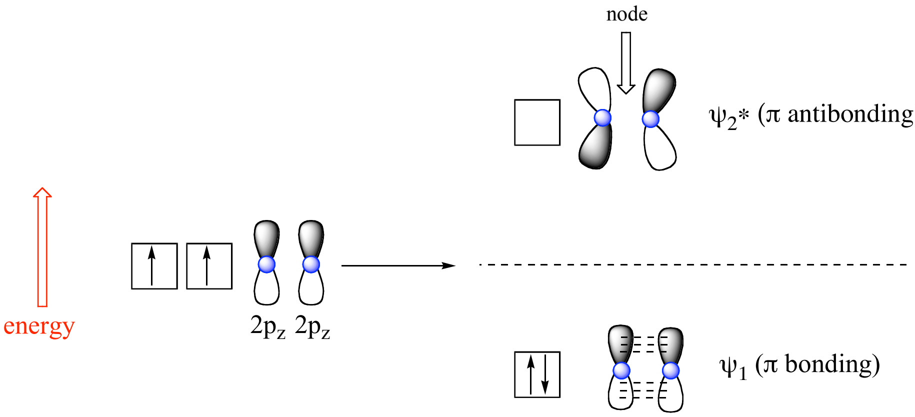
Following the *aufbrau* ('building up') principle, we place the two electrons in the H_2 molecule in the lowest energy orbital, which is the (bonding) σ orbital.

The bonding σ orbital, which holds both electrons in the ground state of the molecule, is egg-shaped, encompassing the two nuclei, and with the highest likelihood of electrons being in the area between the two nuclei. The high-energy, antibonding σ* orbital can be visualized as a pair of droplets, with areas of higher electron density near each nucleus and a 'node', (area of zero electron density) midway between the two nuclei.

Remember that we are thinking here about electron behavior as *wave behavior*. When two separate waves combine, they can do so with what is called constructive interference, where the two amplitudes reinforce one another, or destructive interference, where the two amplitudes cancel one another out. Bonding MO's are the consequence of constructive interference between two atomic orbitals which results in an attractive interaction and an increase in electron density between the nuclei. Antibonding MO's are the consequence of destructive interference which results in a repulsive interaction and a 'canceling out' of electron density between the nuclei (in other words, a node),

2.1B: MO theory and pi bonds - conjugation

The advantage of MO theory becomes more apparent when we think about π bonds, especially in those situations where two or more π bonds are able to interact with one another. Let's first consider the π bond in ethene from an MO theory standpoint (in this example we will be disregarding the various σ bonds, and thinking *only* about the π bond). According to MO theory, the two atomic $2p_z$ orbitals combine to form two **π molecular orbitals**, one a low-energy π bonding orbital and one a high-energy π* antibonding orbital. These are sometimes denoted, in the MO system, using the Greek letter ψ (*psi*) instead of π.

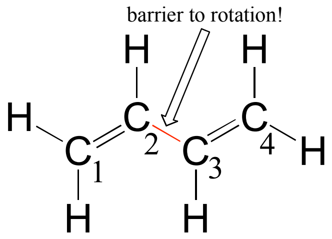


In the bonding ψ_1 orbital, the two shaded lobes of the $2p_z$ orbitals interact constructively with each other, as do the two unshaded lobes (remember, the shading choice represents mathematical (+) and (-) signs for the wavefunction). Therefore, there is increased electron density between the nuclei in the molecular orbital – this is why it is a bonding orbital.

In the higher-energy antibonding ψ_2* orbital, the shaded lobe of one $2p_z$ orbital interacts destructively with the unshaded lobe of the second $2p_z$ orbital, leading to a node between the two nuclei and overall repulsion.

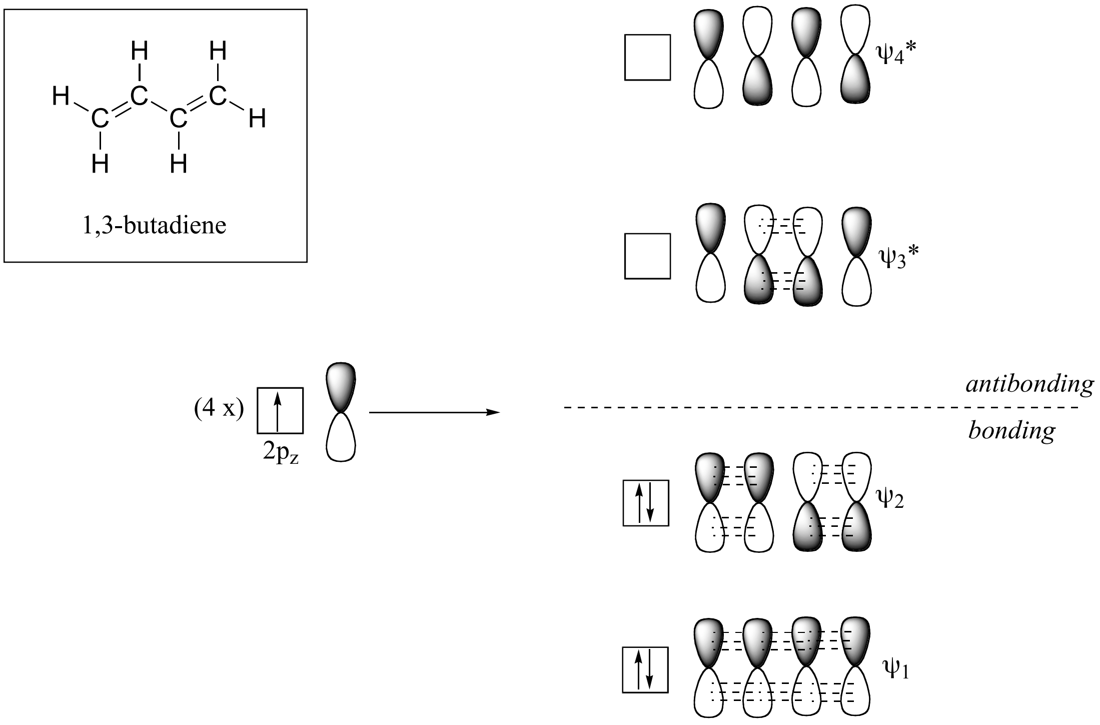
By the *aufbrau* principle, the two electrons from the two atomic orbitals will be paired in the lower-energy ψ_1 orbital when the molecule is in the ground state.

Now, consider the 1,3-butadiene molecule. From valence orbital theory we might expect that the C_2-C_3 bond in this molecule, because it is a σ bond, would be able to rotate freely.



Experimentally, however, it is found that there are significant barriers to rotation about this bond (as well as about the C_1-C_2 and C_3-C_4 double bonds), and that the entire molecule is planar. It is also observed that the C_2-C_3 bond, while longer than the C_1-C_2 and C_3-C_4 double bonds, is significantly shorter than a typical carbon-carbon single bond.

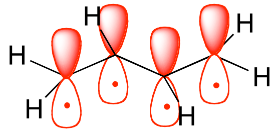
Molecular orbital theory accounts for these observations with the concept of **delocalized π bonds**. In this picture, the four $2p_z$ orbitals are all parallel to each other (and perpendicular to the plane of the σ bonds), and thus there is π overlap not just between C_1 and C_2 and C_3 and C_4, but between C_2 and C_3 as well. The four atomic ($2p_z$) orbitals have combined to form four π molecular orbitals.



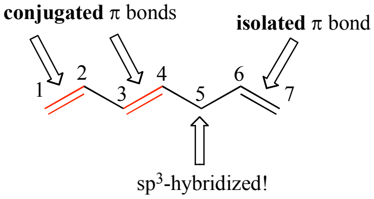
The lowest energy molecular orbital, ψ_1, has zero nodes, and is a bonding MO. Slightly higher in energy, but still lower than the isolated *p* orbitals, is the ψ_2 orbital. This orbital has one node between C_2 and C_3, but is still a bonding orbital due to the two constructive interactions between C_1-C_2 and C_3-C_4. The two higher-energy MO's are denoted ψ_3* and

ψ_4*, and are antibonding. Notice that ψ_3* has two nodes and one constructive interaction, while ψ_3* has three nodes and zero constructive interactions. The energy of both of these antibonding molecular orbitals is higher than that of the $2p_z$ atomic orbitals of which they are composed.

By the *aufbrau* principle, the four electrons from the isolated $2p_z$ atomic orbitals are placed in the bonding ψ_1 and ψ_2 MO's. Because ψ_1 includes constructive interaction between C_2 and C_3, there is a degree, in the 1,3-butadiene molecule, of π-bonding interaction between these two carbons, which accounts for the shorter length and the barrier to rotation. The simple Lewis structure picture of 1,3-butadiene shows the two π bonds as being isolated from one another, with each pair of π electrons 'stuck' in its own π bond. However, molecular orbital theory predicts (accurately) that the four π electrons are to some extent delocalized, or 'spread out', over the whole π system.

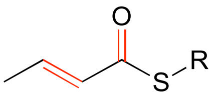


1,3-butadiene is the simplest example of a system of '**conjugated'** π bonds. To be considered conjugated, two or more π bonds must be separated by only one single bond – in other words, there cannot be an intervening sp^3-hybridized carbon, because this would break up the overlapping system of parallel $2p_z$ orbitals. In the compound below, for example, the C_1-C_2 and C_3-C_4 double bonds are conjugated, while the C_6-C_7 double bond is considered to be isolated, due to the effect of the sp^3-hybridized C_5.

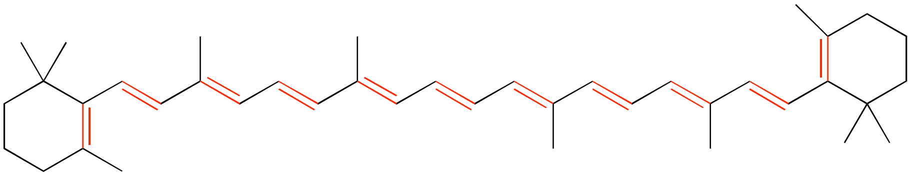


The presence of C_5 makes it impossible for the π electrons of the C_6-C_7 π bond to join the conjugated system on the first four carbons.

Conjugated π systems can involve oxygens and nitrogen atoms as well as carbon. In the metabolism of fat molecules, some of the key reactions involve alkenes that are conjugated to carbonyl groups.

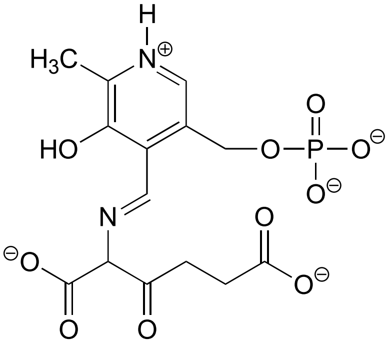


Later on in chapter 4, we will see that MO theory is very useful in explaining why organic molecules that contain extended systems of conjugated π bonds often have distinctive colors. β-carotene, the compound responsible for the orange color of carrots, has an extended system of 11 conjugated π bonds.



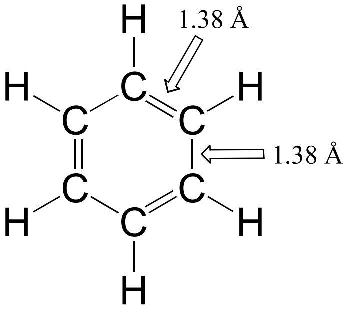
β-carotene

Exercise 2.1: Identify isolated and conjugated π bonds in the compound shown.



2.1C: Aromaticity

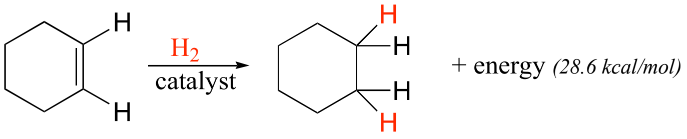
Molecular orbital theory is especially helpful in explaining the unique properties of a class of compounds called aromatics. Benzene, a common organic solvent, is the simplest example of an aromatic compound.



benzene

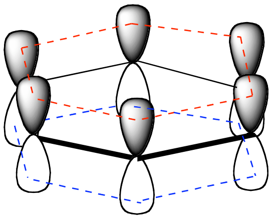
Although it is most often drawn with three double bonds and three single bonds, it is known that all of the carbon-carbon bonds in benzene are exactly the same length - 1.38 Å. This is shorter than a typical carbon-carbon single bond (about 1.54 Å), and slightly longer than a typical carbon-carbon double bond (about 1.34 Å).

In addition, the π bonds in benzene are significantly less reactive than isolated or conjugated π bonds in most alkenes. To illustrate this unique stability, we will make use of the idea of 'heat of hydrogenation'. The carbon-carbon double bond in an alkene can be converted to a single bond through a process called 'catalytic hydrogenation' – essentially adding a molecule of H_2 to the double bond.



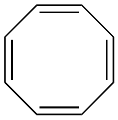
We will learn more about how this process occurs, both in the laboratory and in living cells, later in the text (section 16.5). For now, what is important to understand is that the hydrogenation process is *exothermic*: the alkane is lower in energy than the alkene, so hydrogenating the double bond results in the release of energy in the form of heat. Converting one mole of cyclohexene to cyclohexane, for example, releases 28.6 kilocalories. If the benzene molecule is considered to be a six-membered ring with three isolated double bonds, the heat of hydrogenation should theoretically be three times this value, or 85.8 kcal/mol. The actual heat of hydrogenation of benzene, however, is only 49.8 kcal/mol, or 36 kcal/mol less than what we would expect if using the isolated double bond model. Something about the structure of benzene makes these π bonds especially stable. This 'something' has a name: it is called 'aromaticity'.

What exactly is this 'aromatic' property that makes the π bonds in benzene so much less reactive than those in alkenes? In a large part, the answer to this question lies in the fact that benzene is a *cyclic* molecule in which all of the ring atoms are sp^2-hybridized. This allows the π electrons to be delocalized in molecular orbitals that extend all the way around the ring, above and below the plane of the ring. For this to happen, of course, the ring must be planar – otherwise the $2p_z$ orbitals couldn't overlap properly. Benzene is indeed known to be a flat molecule.



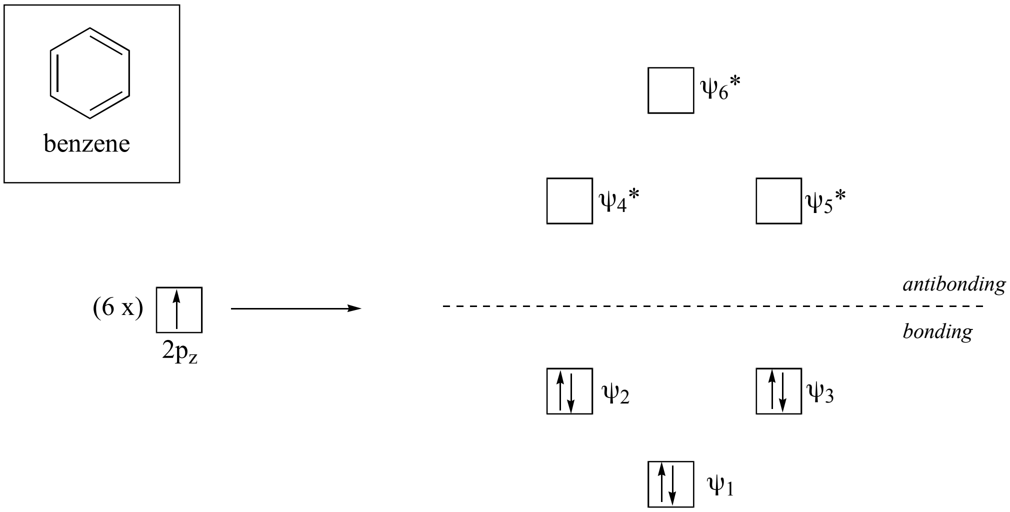
π electrons delocalized around the ring, above and below the plane

Do all cyclic molecules with alternating single and double bonds have this same aromatic stability? Quite simply, the answer is 'no'. The eight-membered cyclooctatetraene ring shown below is *not* flat, and its π bonds are much more reactive than those of benzene.



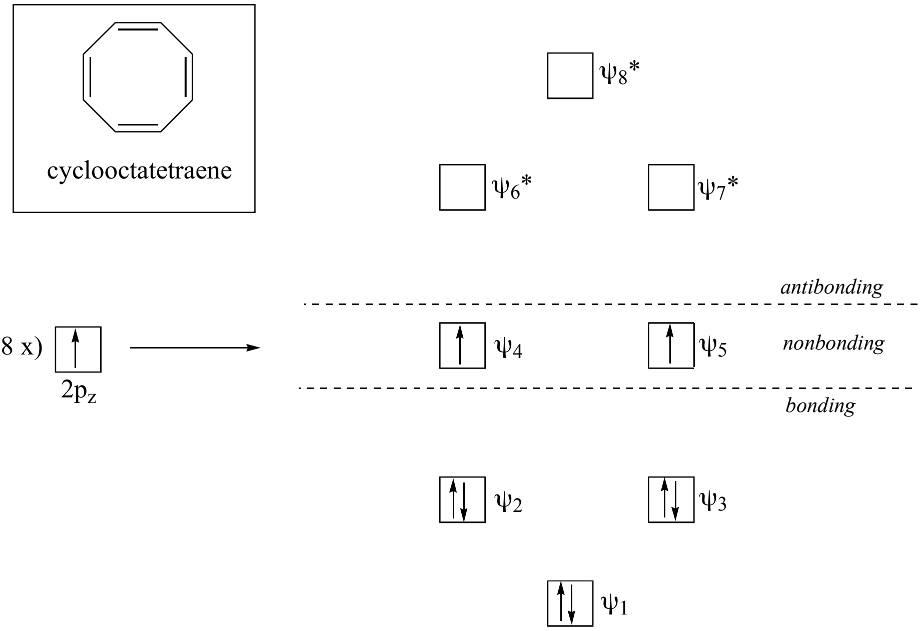
cyclooctatetraene

Clearly it takes something more to be aromatic, and this can best be explained with molecular orbital theory. Let's look at an energy diagram for the molecular orbitals containing the π electrons in benzene.



Quantum mechanical calculations conclude that the six molecular orbitals in benzene, formed from six atomic $2p_z$ orbitals, occupy four separate energy levels. ψ_1 and ψ_6* have unique energy levels, while the ψ_2 - ψ_3 and ψ_4*-ψ_5* pairs are **degenerate** (more than one orbital at the same energy level). When we use the *aufbrau* principle to fill up these orbitals with the six π electrons in benzene, we see that the bonding orbitals are completely filled, and the antibonding orbitals are empty. This gives us a good clue to the source of the special stability of benzene: a full set of bonding MO's is similar in many ways to the 'full shell' of electrons possessed by the very stable noble gases like helium, neon, and argon.

Now, let's do the same thing for cyclooctatetraene, which we have already learned is not aromatic.



The result of molecular orbital calculations tells us that the lowest and highest energy MOs (ψ_1 and ψ_8*) have unique energy levels, while the other six come in degenerate pairs. Notice that ψ_4 and ψ_5 are at the same energy level as the isolated $2p_z$ atomic orbitals: these are therefore neither bonding nor antibonding, rather they are referred to as **nonbonding MOs**. Filling up the MOs with the eight π electrons in the molecule, we find that the last two electrons are unpaired and fall into the two degenerate nonbonding orbitals. Because we don't have a perfect filled shell of bonding MOs, our molecule is not aromatic. As a consequence, each of the double bonds in cyclooctatetraene acts more like an *isolated* double bond.

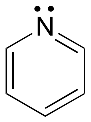
Here, then, are the conditions that must be satisfied for a molecule to be considered aromatic:

1) It must have a cyclic structure.
2) The ring must be planar.
3) Each atom in the ring must be sp^2-hybridized, so that π electrons can be delocalized around the ring.
4) The number of π electrons in the ring must be such that, in the ground state of the molecule, all bonding MOs are completely filled, and all nonbonding and antibonding MOs are completely empty.

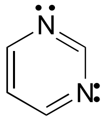
It turns out that, in order to satisfy condition #4, the ring must contain a specific number of π electrons. The set of possible numbers is quite easy to remember - the rule is simply $4n+2$, where n is any positive integer (this is known as the **Hückel rule**, named after Erich Hückel, a German scientist who studied aromatic compounds in the 1930's). Thus, if $n = 0$, the first Hückel number is $(4 \times 0) + 2$, or 2. If $n = 1$, the Hückel number is $(4 \times 1) + 2$, or 6 (the Hückel number for benzene). The series continues with 10, 14, 18, 22, and so on. Cyclooctatetraene has eight π electrons, which is *not* a Hückel number. Because 6 is such a common Hückel number, chemists often use the term **'aromatic sextet'**.

Benzene is best visualized as a planar ring made up of carbon-carbon σ bonds, with two 'donut-like' rings of fully delocalized π electron density above and below the plane of the ring (the fact that there is a ring of π electron density on *both* sides of the molecule stems from the fact that the overlapping *p* orbitals have two lobes, and the electron density is located in both). This general picture is valid not just for benzene but for all other aromatic structures as well.

Let's look at some different aromatic compounds other than benzene. Pyridine and pyrimidine both fulfill all of the criteria for aromaticity.



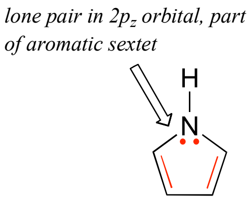
pyridine



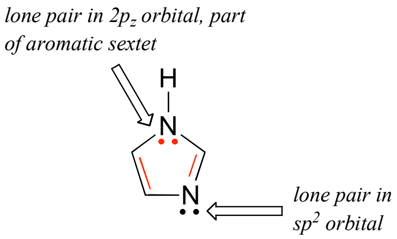
pyrimidine

In both of these molecules, the nitrogen atoms are sp^2 hybridized, with the lone pair occupying an sp^2 orbital and therefore not counted among the aromatic sextet. The Hückel number for both pyridine and pyrimidine is six.

Rings do not necessarily need to be 6-membered in order to have six π electrons. Pyrrole and imidizole, for example, are both aromatic 5-membered rings with six π electrons.



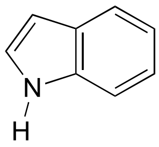
pyrrole

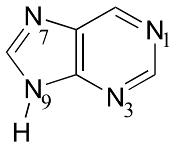

lone pair in $2p_z$ orbital, part of aromatic sextet
H
N
N
lone pair in sp^2 orbital

imidazole

The nitrogen atoms in both of these molecules are sp^2-hybridized (as they must be for the rings to be aromatic). In pyrrole, the lone pair can be thought of as occupying a $2p_z$ orbital, and thus both of these electrons contribute to the aromatic π system. In imidazole, one lone pair occupies a $2p_z$ orbital and is part of the aromatic sextet, while the second occupies one of the sp^2 orbitals and is not part of the sextet.

Molecules with more then one ring can also fulfill the Hückel criteria, and often have many of the same properties as monocyclic aromatic compounds, including a planar structure. Indole (a functional group in the amino acid tryptophan) and purine (a functional group in guanine and adenine DNA/RNA bases) both have a total of ten π electrons delocalized around two rings.


indole

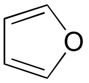


purine

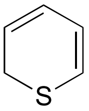
The nitrogen in indole and the N_9 nitrogen in purine both contribute a pair of electrons to the π system. The N_1, N_3, and N_7 nitrogens of purine, in contrast, hold their lone pair in sp^2 orbitals, outside of the aromatic system.

Exercise 2.2: Are the following molecules likely to be aromatic? Explain, using Huckel's criteria. and orbital drawings. Hint: Ions can also be aromatic!

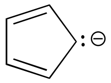
a)



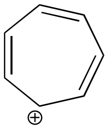
b)



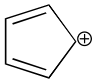
c)



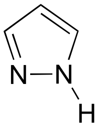
d)



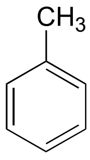
e)

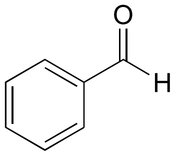


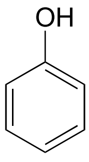
f)



Up to now we have been talking about molecules in which the entire structure makes up an aromatic system. However, in organic chemistry we will more often encounter examples of molecules which have both aromatic *and* nonaromatic parts. Toluene, a common organic solvent (which is much safer to use than benzene) is simply a benzene ring with a methyl substituent. Benzyaldehyde is benzene with an aldehyde substituent, and phenol is benzene with a hydroxyl substituent.

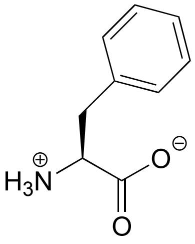

toluene


benzaldehyde

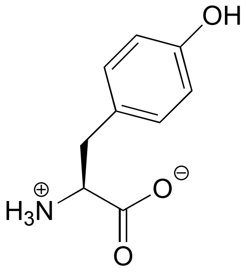

phenol

In these 'substituted benzene' compounds, the entire molecule is not aromatic, just the benzene ring part.

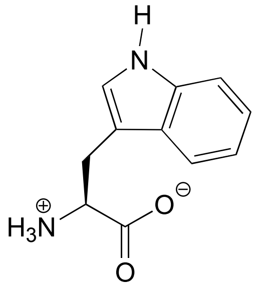
When a benzene ring is part of a larger molecule, it is called a 'phenyl' group. The amino acid phenylalanine, for example, contains a phenyl group. The amino acids tyrosine, tryptophan, and histidine contain phenol, indole, and imidazole groups, respectively.



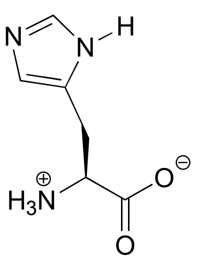
phenylalanine



tyrosine

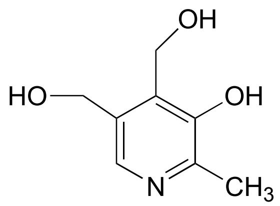


tryptophan



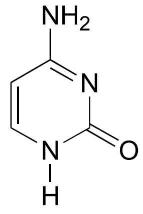
histidine

Pyridoxine, commonly known as vitamin B_6, is a substituted pyridine.

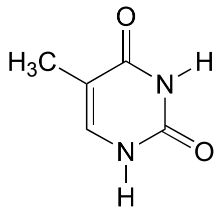


pyridoxine
(vitamin B_6)

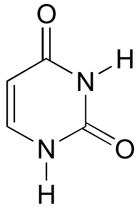
The DNA and RNA bases are based on pyrimidine (cytosine, thymine, and uracil) and purine (adenine and guanine).



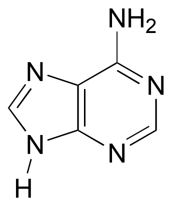
cytosine



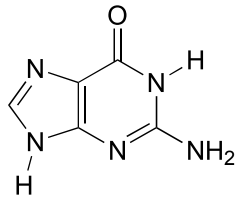
thymine



uracil



adenine



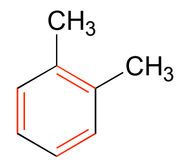
guanine

The flat, aromatic structure of these bases plays a critical role in the overall structure and function of DNA and RNA.

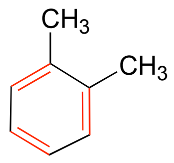
Section 2.2: Resonance

2.2A: The meaning of resonance contributors: benzene and its derivatives

If we were to draw the structure of an aromatic molecule such as 1,2-dimethylbenzene, there are two ways that we could draw the double bonds:



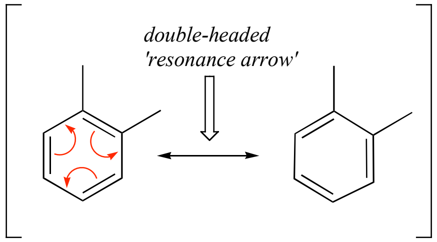
is this correct. . .



or is this?

Which way is correct? There are two simple answers to this question: 'both' and 'neither one'. Both ways of drawing the molecule are equally acceptable approximations of the bonding picture for the molecule, but neither one, by itself, is an accurate picture of the delocalize π bonds. The two alternative drawings, however, *when considered together*, give a much more accurate picture than either one on its own. This is because they imply, together, that the carbon-carbon bonds are not double bonds, not single bonds, but about halfway in between.

These two drawings are an example of what is referred to in organic chemistry as **resonance contributors**: two or more different Lewis structures depicting the same molecule or ion that, when considered together, do a better job of approximating delocalized π bonding than any single structure. By convention, resonance contributors are linked by a double-headed arrow, and are sometimes enclosed by brackets:



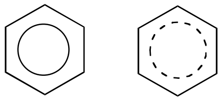
resonance contributor A

resonance contributor B

In order to make it easier to visualize the difference between two resonance contributors, small, curved arrows are often used. Each of these arrows depicts the 'movement' of two π electrons. A few chapters from now when we begin to study organic reactions - a process in which electron density shifts and covalent bonds between atoms break and form - this 'curved arrow notation' will become extremely important in depicting electron movement. In the drawing of resonance contributors, however, this electron 'movement' occurs only in our minds, as we try to visualize delocalized π bonds. Nevertheless, use of the curved arrow notation is an essential skill that you will need to develop in drawing resonance contributors.

The depiction of benzene using the two resonance contributors A and B in the figure above does *not* imply that the molecule at one moment looks like structure A, then at the next moment shifts to look like structure B. Rather, at all moments, the molecule is a combination, or **resonance hybrid** of both A and B.

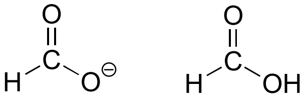
Usually, benzene and substituted derivatives of benzene are depicted with only one resonance contributor, and it is assumed that the reader understands that resonance delocalization is implied. This is the convention that will be used for the most part in this book. In other books or articles, you may sometimes see benzene or a phenyl group drawn with a circle inside the hexagon, either solid or dashed, as a way of depicting the resonance delocalization.



other ways of depicting benzene

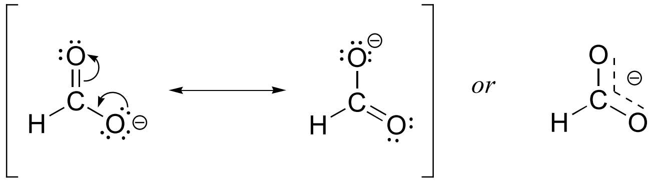
2.2B: Resonance contributors for the carboxylate group

The convention of drawing two or more resonance contributors to approximate a single structure may seem a bit clumsy to you at this point, but as you gain experience you will see that the practice is actually very useful when discussing the manner in which many functional groups react. Let's next consider the carboxylate ion (the conjugate base of a carboxylic acid). As our example, we will use formate, the simplest possible carboxylate-containing molecule. The conjugate acid of formate is formic acid, which causes the painful sting you felt if you have ever been bitten by an ant.

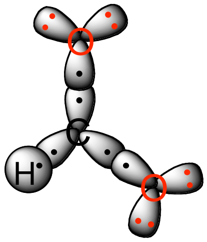


formate formic acid

Usually, you will see carboxylate groups drawn with one carbon-oxygen double bond and one carbon-oxygen single bond, with a negative formal charge located on the single-bonded oxygen. In actuality, however, the two carbon-oxygen bonds are the same length, and although there is indeed an overall negative formal charge on the group, it is shared equally between the two oxygens. Therefore, the carboxylate can be more accurately depicted by a *pair* of resonance contributors. Alternatively, a single structure can be used, with a dashed line depicting the resonance-delocalized π bond and the negative charge located in between the two oxygens.

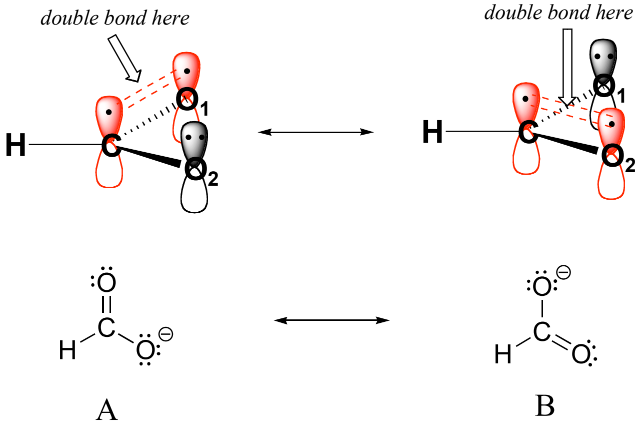


Let’s see if we can correlate these drawing conventions to a valence bond theory picture of the bonding in a carboxylate group. We know that the carbon must be sp^2-hybridized, (the bond angles are close to 120°, and the molecule is planar), and we will treat both oxygens as being sp^2-hybridized as well. Both carbon-oxygen σ bonds, then, are formed from the overlap of carbon sp^2 orbitals and oxygen sp^2 orbitals.



the σ-bonding framework of formate

In addition, the carbon and both oxygens each have an unhybridized $2p_z$ orbital situated perpendicular to the plane of the σ bonds. These three $2p_z$ orbitals are parallel to each other, and can overlap in a side-by-side fashion to form a delocalized π bond.



Resonance contributor A shows oxygen #1 sharing a pair of electrons with carbon in a π bond, and oxygen #2 holding a lone pair of electrons in its $2p_z$ orbital. Resonance contributor B, on the other hand, shows oxygen #2 participating in the π bond with carbon, and oxygen #1 holding a lone pair in its $2p_z$ orbital. Overall, the situation is one of *three parallel, overlapping $2p_z$ orbitals sharing four delocalized π electrons*. Because there is one more electron than there are $2p_z$ orbitals, the system has an overall charge of –1. This is the kind of 3D picture that resonance contributors are used to approximate, and once you get some practice you should be able to quickly visualize overlapping $2p_z$

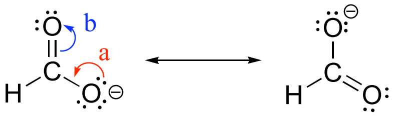
orbitals and delocalized π electrons whenever you see resonance structures being used. In this text, carboxylate groups will usually be drawn showing only one resonance contributor for the sake of simplicity, but you should always keep in mind that the two C-O bonds are equal, and that the negative charge is delocalized to both oxygens.

2.2C: Rules for drawing resonance structures

As you work on learning how to draw and interpret resonance structures, there are a few basic rules that you should keep in mind in order to avoid drawing nonsensical structures. All of these rules make perfect sense as long as you keep in mind that resonance contributors are merely a human-invented convention for depicting the delocalization of electrons in overlapping *p* orbitals.

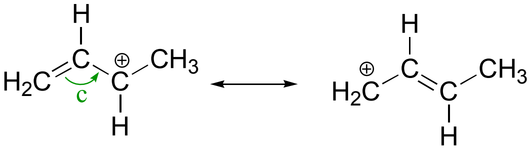
> 1) When you see two different resonance contributors, you are *not* seeing a chemical reaction! Rather, you are seeing the exact same molecule or ion depicted in two different ways.
>
> 2) Resonance contributors involve the 'imaginary movement' of π-bonded electrons and of lone pairs that are adjacent to (*i.e.* conjugated to) π bonds. You can *never* shift the location of electrons in σ bonds – if you show a σ bond forming or breaking, you are showing a chemical reaction taking place (see rule #1).
>
> 3) All resonance contributors for a molecule or ion must have the same *net* charge – if not, re-check your formal charge calculations.
>
> 4) Be very careful in your electron accounting so as not to break the octet rule. Never show "electron movement' arrows that would lead to a situation where a second-row element has more than eight electrons. Sometimes, however, resonance contributors can be drawn in which a second-row atom has only six electrons (see resonance rule #5 below).

To expand a bit on rule #4, there are really only three things we can do with curved arrows when drawing resonance structures. First, we can take a lone pair on an atom and put those two electrons into a π bond on the same atom (arrow 'a' below). Second, we can take the two electrons in a π bond and shift them up into a lone pair (arrow 'b' below).



Notice that the electron movement shown by arrow 'b' allows us to avoid placing ten electrons around the carbon.

Third, we can take the two electrons in a π bond and shift them over by one bond. Arrow 'c' below illustrates this third allowable possibility.

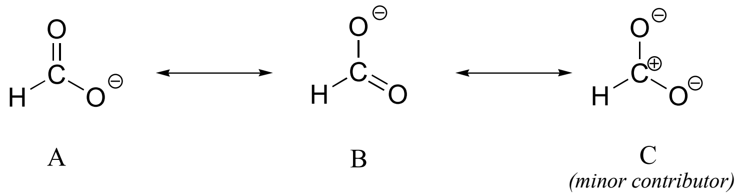


Notice that we do not exceed the octet rule on any carbons when we move the π electrons, because the positively charged carbon only has six electrons around it to start off with.

When drawing resonance structures, it will help you to keep track of electrons and bonds if you carefully draw arrows to show how you are moving electrons to go from one resonance contributor to another. Always be very careful that your arrows do ***only*** the three types of electron movement described above, and that you ***never*** exceed the octet rule on a second-row element.

2.2D: Major vs minor resonance contributors: four more rules to follow

Different resonance contributors do not always make the same contribution to the overall structure of the hybrid. If one resonance structure is more stable (lower in energy) than another, then the first will come closer to depicting the 'real' (hybrid) structure than the second. In the case of carboxylates, contributors A and B are equivalent to each other in terms of their relative stability and therefore their relative contribution to the hybrid structure. However, there is a third resonance contributor 'C' that we have not considered yet, in which the carbon bears a positive formal charge and both oxygens are single-bonded and bear negative charges.



Structure C is relatively less stable, and therefore makes a less important contribution to the overall structure of the hybrid relative to A and B.

How do we know that structure C is the less stable, and thus the 'minor' contributor? There are four basic rules which you need to learn in order to evaluate the relative stability of different resonance contributors. We will number them 5-8 so that they may be added to in the 'rules for resonance' list from section 2.2C.

> 5) The carbon in contributor C does not have an octet – in general, resonance contributors in which a carbon does not fulfill the octet rule are relatively less important.
>
> 6) In structure C, a separation of charge has been introduced that is not present in A or B. In general, resonance contributors in which there is a greater separation of charge are relatively less important.

> 7) In structure C, there are no double bonds. In general, a resonance structure with a lower number of multiple bonds is relatively less important.

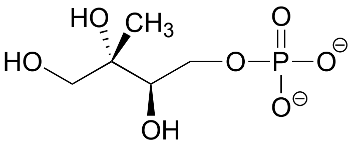
There is one more important rule that does not apply to this particular example, but we will list it here in the interest of completeness (we will soon see an actual example).

> 8) The resonance contributor in which a negative formal charge is located on a more electronegative atom (such as oxygen or nitrogen) is more stable than one in which the negative charge is located on a less electronegative atom (such as carbon).

When discussing other examples of resonance contributors, we will often see cases where one form is less stable – but these minor resonance contributors can still be very relevant in explaining properties of structure or reactivity, and should not be disregarded.

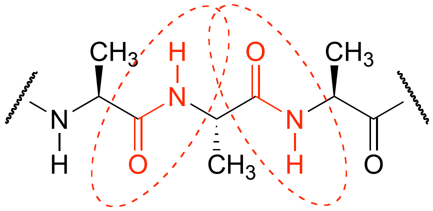
Exercise 2.3: Draw a second resonance structure for acetone (IUPAC name 2-propanone). Explain why this is a minor (less stable) contributor.

Exercise 2.4: Draw four additional resonance contributors for the molecule below. Label each one as major or minor.



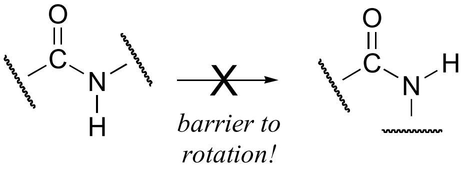
2.2E: More examples of resonance contributors: peptide bonds, enolates, and carbocations

A consideration of resonance contributors is crucial to any discussion of the amide functional group. One of the most important examples of amide groups in nature is the 'peptide bond' that links amino acids to form polypeptides and proteins.

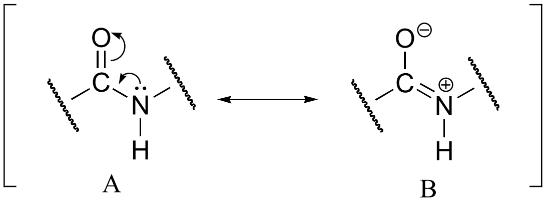


peptide bonds in a polypeptide

Critical to the structure of proteins is the fact that, although it is conventionally drawn as a single bond, the C-N bond in a peptide linkage has a significant barrier to rotation, almost as if it were a double bond.



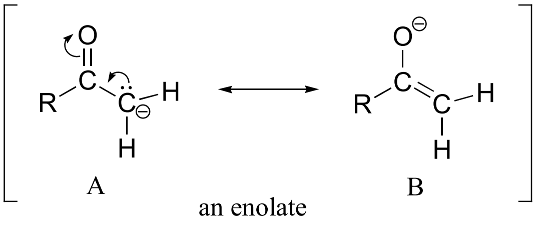
This, along with the observation that the bonding around the peptide nitrogen has trigonal planar geometry, strongly suggests that the nitrogen is sp^2-hybridized. An important resonance contributor has a C=N double bond and a C-O single bond, with a separation of charge between the oxygen and the nitrogen.



Although B is a minor contributor due to the separation of charges, it is still very relevant in terms of peptide and protein structure – our proteins would simply not fold up properly if there was free rotation about the peptide C-N bond.

Exercise 2.5: Draw two pictures showing the unhybridized *p* orbitals and the location of π electrons in an amide group. One picture should represent the major resonance contributor A, the other the minor contributor B (from the figure above).

'Enolate ions' are extremely important species in organic chemistry – if you look ahead, you will see that two whole chapters are devoted to reactions in which enolates play a role! In an enolate, a negatively-charged carbon (carbanion) is adjacent to a carbonyl group (this could be on an aldehyde, a ketone, or a carboxylic acid derivative).

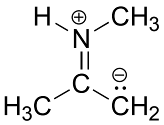


an enolate

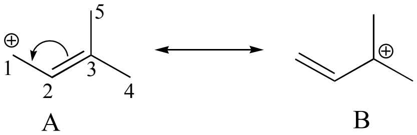
A second resonance contributor B puts the negative charge on the oxygen, and shows a carbon-carbon double bond.

Which of the two resonance contributors is more important? Consider now rule #8 in the list presented earlier. We have not yet discussed electronegativity (that is the subject of section 2.3), but you probably remember from your general chemistry class that oxygen is more electronegative than carbon. Therefore, contributor B is the more stable of the two. However, as we will see in chapters 13 and 14, enolate ions usually react in ways that suggest that the carbon bears a substantial degree of negative charge, so contributor A is still extremely relevant.

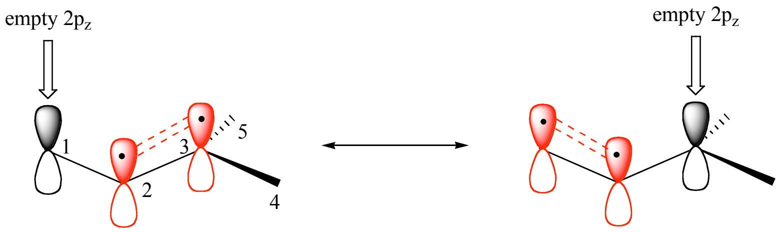
Exercise 2.6: Draw the structure of a more stable resonance contributor for the molecule shown below.



Carbocations such as the one depicted below by two resonance contributors are important in the synthesis of cholesterol and many other related biological molecules.



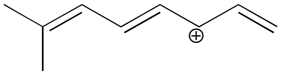
In this case, three parallel $2p_z$ orbitals are sharing two π electrons: because there is one fewer π electron than there are $2p_z$ orbitals, there is a positive charge overall.



For reasons that will be discussed in chapter 7, contributor B is the more stable of the two. The use of these two resonance contributors expresses the idea that both carbon #1 and carbon #3 bear some degree of positive charge, and could react accordingly.

Exercise 2.7:

a) Draw three more resonance contributors for the carbocation shown.

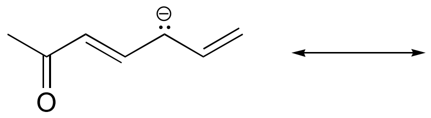


b) Fill in the blanks: the conjugated π system in this carbocation is composed of ______ p orbitals containing ________ delocalized π electrons.

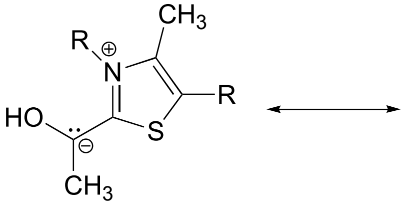
Exercise 2.8:

Draw the most stable resonance contributor for each of the anions below.

a)



b)



c) Fill in the blanks: the conjugated π system in structure a) is composed of ______ p orbitals containing ________ delocalized π electrons.

A final word of advice: becoming adept at drawing resonance contributors - and understanding the overall concept of resonance delocalization - is one of the most important jobs that you will have as a beginning student of organic chemistry. If you work hard now to gain a firm grasp of these ideas, you will have come a long way toward understanding much of what follows in your organic chemistry course. Conversely, if you fail to come to grips with these concepts now, a lot of what you see later in the course will seem mysterious and incomprehensible, and you will be in for a rough ride, to say the least. Keep working problems, keep asking questions, keep pounding at it until you get it!

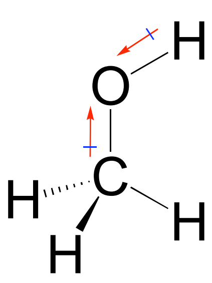
Section 2.3: Non-covalent interactions

Until now we have been focusing on understanding the covalent bonds that hold individual molecules together. We turn next to a brief review on the subject of non-covalent interactions *between* molecules, or between different functional groups within a single molecule.

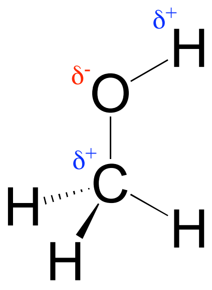
2.3A: Dipoles

To understand the nature of noncovalent interactions, we first must return to covalent bonds and delve into the subject of dipoles. Many of the covalent bonds that we saw in section 1.5 – between two carbons, for example, or between a carbon and a hydrogen – involve the approximately equal sharing of electrons between the two atoms in the bond. This is because, in these examples, the two atoms have approximately the same **electronegativity**. Recall from general chemistry that electronegativity refers to " the power of an atom in a molecule to attract electrons to itself" (this is the definition offered by Linus Pauling, the eminent 20th-century American chemist who was primarily responsible for developing many of the bonding concepts that we have been learning).

Quite often, however, we deal in organic chemistry with covalent bonds between two atoms with very different negativities, and in these cases the sharing of electrons is not equal: the more electronegative atom pulls the two electrons closer to itself. In the carbon-oxygen bond of an alcohol, for example, the two electrons in the σ bond are closer to the oxygen than they are to the carbon, because oxygen is significantly more electronegative than carbon. The same is true for the oxygen-hydrogen bond, as hydrogen is slightly less electronegative than carbon, and much less electronegative than oxygen.



bond dipole arrows



partial charge notation

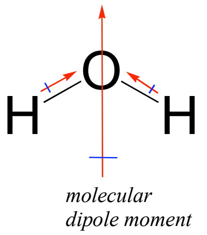
The result of this unequal sharing is what we call a **bond dipole**, or a **polar covalent bond**. A polar covalent bond has both negative and positive ends, or poles, where electron density is lower (the positive pole) and higher (the negative pole). The difference in electron density can be expressed using the Greek letter ∂ (delta) to express 'partial positive' and 'partial negative' charge on the atoms. 'Dipole arrows', with a positive sign on the tail, are also used to indicated the direction of higher electron density.

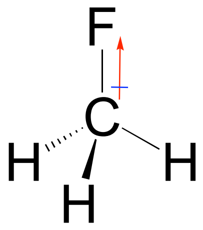
The degree of polarity depends on the difference in electronegativity between the two atoms in the bond. Electronegativity is a periodic trend: it increases going from left to right across a row of the periodic table of the elements, and also increases as we move up a column. Therefore, oxygen is more electronegative than nitrogen, which is in turn more electronegative than carbon. Oxygen is also more electronegative than sulfur. Fluorine, in the top right corner of the periodic table, is the most electronegative of the elements, a property which we will see in later chapters can be exploited by researchers who are trying to work out the details of how organic reactions occur.

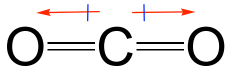
Exercise 2.9: Using the ideas that you have learned about atomic orbitals, rationalize the periodic trends in electronegativity.

Most molecules contain both polar and nonpolar covalent bonds. Depending on the location of polar bonds and bonding geometry, molecules may posses an overall **dipole**

moment. Water, as you probably recall, has a dipole moment that results from the combined dipoles of its two oxygen-hydrogen bonds. Fluoromethane also has an overall dipole moment.



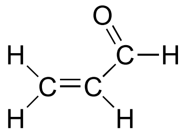




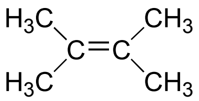
Carbon dioxide, however, has polar bonds that pull in opposite directions, meaning that there is no overall molecular dipole moment.

Exercise 2.10: Which of the molecules below have overall dipole moments?

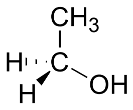
a)



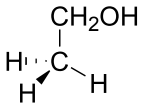
b)



c)

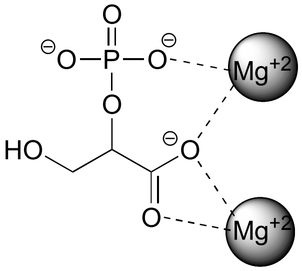


d)

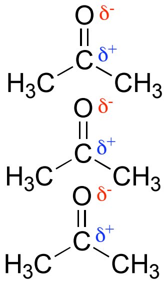


2.3B: Ion-ion, dipole-dipole and ion-dipole interactions

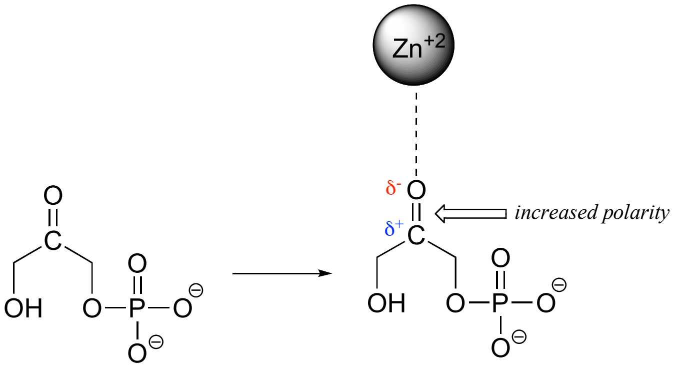
Arguably the simplest kind of non-covalent interaction is one that we have already discussed: the ionic bond formed by the electrostatic attraction between positively and negatively charged species. One of the most common types of ion-ion interaction in biological chemistry is that between a magnesium ion (Mg^{+2}) and a carboxylate or phosphate group. 2-phosphoglycerate, an intermediate in the breakdown of glucose, provides a good example in the way that it interacts with two Mg^{+2} ions while binding to an enzyme called enolase.



Polar molecules – those with an overall dipole moment, such as acetone – will align themselves in such a way as to allow their respective positive and negative poles to interact with each other. This is called a '**dipole-dipole interaction**.



When a charged species (an ion) interacts favorably with a polar molecule or functional group, the result is called an **ion-dipole interaction**. A common example of ion-dipole interaction in biological organic chemistry is that between a metal cation, most often Mg^{+2} or Zn^{+2}, and the partially negative oxygen of a carbonyl.



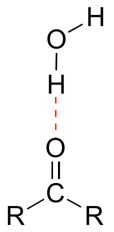
Because the metal cation is very electronegative, this interaction has the effect of pulling electron density in the carbonyl double bond even further toward the oxygen side, increasing the partial positive charge on carbon. As we shall later, this has important implications in terms of the reactivity of carbonyl groups.

2.3C: van der Waals forces

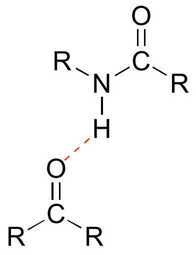
Nonpolar molecules such as hydrocarbons also are subject to relatively weak but still significant attractive intermolecular forces. **Van der Waals forces** (also called **London dispersion forces**, **hydrophobic inter**actions, or **nonpolar interactions**) result from the constantly shifting electron density in any molecule. Even a nonpolar molecule will, at any given moment, have a weak, short-lived dipole. This transient dipole will induce a neighboring nonpolar molecule to develop a corresponding transient dipole of its own, with the end result that a weak dipole-dipole force is formed. These van der Waals forces are relatively weak by nature, but are constantly forming and dissipating among closely-packed nonpolar molecules, and when added up the cumulative effect can become significant. It is through van der Waals interactions that the nonpolar hydrocarbon regions of fat molecules hold together to form the lipid bilayer in cells and subcellular organelles (we will discuss this topic in more detail in section 2.4B)

2.3D: Hydrogen bonds

Hydrogen bonds are an extremely important form of noncovalent interaction. Hydrogen bonds result from the interaction between a hydrogen that is bound to a electronegative heteroatom – specifically a nitrogen or oxygen, or sometimes a fluorine – and a lone pair on a second nitrogen or oxygen. Because a hydrogen is just a single proton with no inner shell (non-valence) electrons, when it loses electron density in a polar bond it essentially becomes an approximation of a 'naked' proton, capable of forming a strong interaction with a lone pair on a neighboring oxygen or nitrogen atom.



hydrogen bonding between
a ketone (acceptor) and
water (donor)



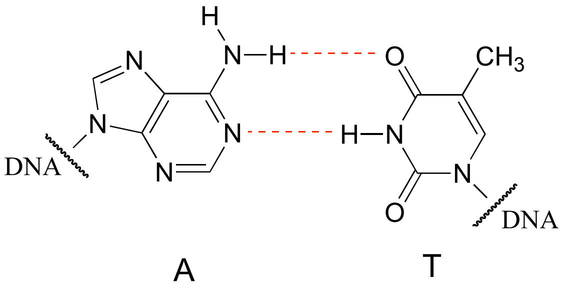
hydrogen bonding between
a ketone (acceptor) and an
amide (donor)

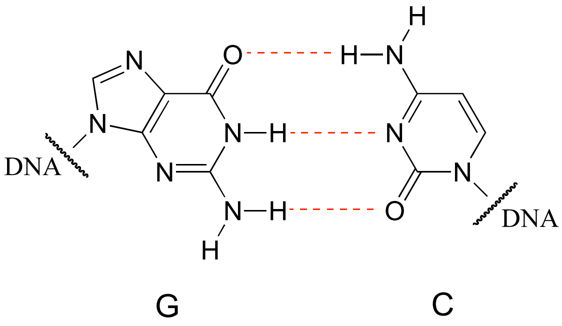
Hydrogen bonds are usually depicted with dotted lines in chemical structures. A group that provides a proton to a hydrogen bond is said to be acting as a **hydrogen bond donor**. A group that provides an oxygen or nitrogen lone pair is said to be acting as a **hydrogen bond acceptor**. Many common organic functional groups can participate in the formation of hydrogen bonds, either as donors, acceptor, or both. A water molecule, for example, can be both a hydrogen bond donor and acceptor. A carbonyl, as it lacks a hydrogen bound to an oxygen or nitrogen, can only act as a hydrogen bond acceptor.

Exercise 2.11: Draw figures that show the hydrogen bonds described below.

a) A hydrogen bond between methanol (donor) and water (acceptor).
b) A hydrogen bond between methanol (acceptor) and water (donor).
c) Two possible hydrogen bonds between methyl acetate and methylamine.

Hydrogen bonds are the strongest type of noncovalent interaction, quite a bit stronger than a dipole-dipole interaction, but also much weaker than a covalent bond. The strength of hydrogen bonds has enormous implications in biology. Genetic inheritance, for example, is based on a very specific hydrogen bonding arrangement between DNA bases: A 'base-pairs' with T, while G 'base-pairs' with C.





Section 2.4: The relationship between noncovalent interactions and the physical properties of organic compounds

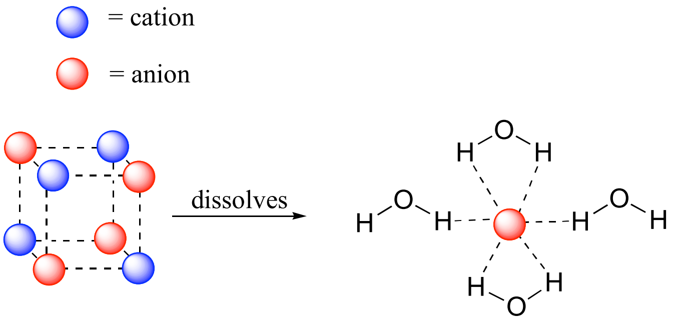
An understanding of bond dipoles and the various types of noncovalent intermolecular forces allows us to explain, on a molecular level, many observable physical properties of organic compounds. In this section, we will concentrate on solubility, melting point, and boiling point.

2.4A: Solubility

Virtually all of the organic chemistry that you will see in this course takes place in the solution phase. In the organic laboratory, reactions are often run in nonpolar or slightly polar solvents such as toluene (methylbenzene), hexane, dichloromethane, or diethylether. In recent years, much effort has been made to adapt reaction conditions to allow for the use of 'greener' (in other words, more environmentally friendly) solvents such as water or ethanol, which are polar and capable of hydrogen bonding. In organic reactions that occur in the cytosolic region of a cell, the solvent is of course water. It is critical for any organic chemist to understand the factors which are involved in the solubility of different molecules in different solvents.

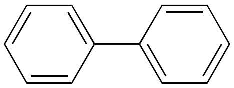
You probably remember the rule you learned in general chemistry regarding solubility: 'like dissolves like' (and even before you took any chemistry at all, you probably observed at some point in your life that oil does not mix with water). Let's revisit this old rule, and put our knowledge of covalent and noncovalent bonding to work.

Imagine that you have a flask filled with water, and a selection of substances that you will test to see how well they dissolve in the water. The first substance is table salt, or sodium chloride. As you would almost certainly predict, especially if you've ever inadvertently taken a mouthful of water while swimming in the ocean, this ionic compound dissolves readily in water. Why? Because water, as a very polar molecule, is able to form many ion-dipole interactions with both the sodium cation and the chloride anion, the energy from which is more than enough to make up for energy required to break up the ion-ion interactions in the salt crystal and some water-water hydrogen bonds.



The end result, then, is that in place of sodium chloride crystals, we have individual sodium cations and chloride anions surrounded by water molecules – the salt is now *in solution*. Charged species as a rule dissolve readily in water: in other words, they are very **hydrophilic** (water-loving).

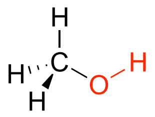
Now, we'll try a compound called biphenyl, which, like sodium chloride, is a colorless crystalline substance (the two compounds are readily distinguishable by sight, however – the crystals look quite different).



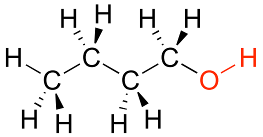
biphenyl

Biphenyl does not dissolve at all in water. Why is this? Because it is a very non-polar molecule, with only carbon-carbon and carbon-hydrogen bonds. It is able to bond to itself very well through nonpolar van der Waals interactions, but it is not able to form significant attractive interactions with the very polar solvent molecules. Thus, the energetic cost of breaking up the biphenyl-to-biphenyl interactions in the solid is high, and very little is gained in terms of new biphenyl-water interactions. Water is a terrible solvent for nonpolar hydrocarbon molecules: they are very **hydrophobic** (water-hating).

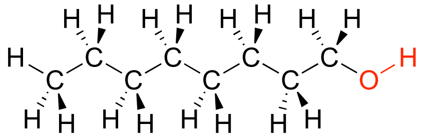
Next, you try a series of increasingly large alcohol compounds, starting with methanol (1 carbon) and ending with octanol (8 carbons).



methanol
very soluble in water



butanol
slightly soluble in water

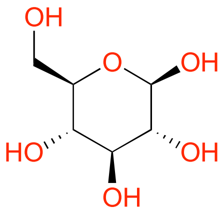


octanol
very insoluble in water

You find that the smaller alcohols - methanol, ethanol, and propanol - dissolve easily in water. This is because the water is able to form hydrogen bonds with the hydroxyl group in these molecules, and the combined energy of formation of these water-alcohol hydrogen bonds is more than enough to make up for the energy that is lost when the alcohol-alcohol hydrogen bonds are broken up. When you try butanol, however, you begin to notice that, as you add more and more to the water, it starts to form its own layer on top of the water.

The longer-chain alcohols - pentanol, hexanol, heptanol, and octanol - are increasingly non-soluble. What is happening here? Clearly, the same favorable water-alcohol hydrogen bonds are still possible with these larger alcohols. The difference, of course, is that the larger alcohols have larger nonpolar, hydrophobic regions in addition to their hydrophilic hydroxyl group. At about four or five carbons, the hydrophobic effect begins to overcome the hydrophilic effect, and water solubility is lost.

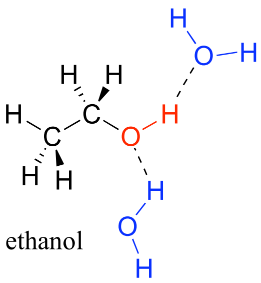
Now, try dissolving glucose in the water – even though it has six carbons just like hexanol, it also has five hydrogen-bonding, hydrophilic hydroxyl groups in addition to a sixth oxygen that is capable of being a hydrogen bond acceptor.



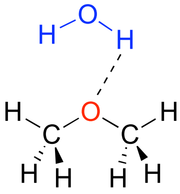
glucose

We have tipped the scales to the hydrophilic side, and we find that glucose is quite soluble in water.

We saw that ethanol was very water-soluble (if it were not, drinking beer or vodka would be rather inconvenient!) How about dimethyl ether, which is a constitutional isomer of ethanol but with an ether rather than an alcohol functional group? We find that diethyl ether is much less soluble in water. Is it capable of forming hydrogen bonds with water? Yes, in fact, it is –the ether oxygen can act as a hydrogen-bond acceptor. The difference between the ether group and the alcohol group, however, is that the alcohol group is both a hydrogen bond donor *and* acceptor.



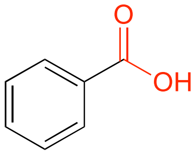
ethanol



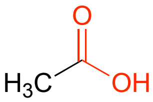
dimethylether

The result is that the alcohol is able to form more energetically favorable interactions with the solvent compared to the ether, and the alcohol is therefore more soluble.

Here is another easy experiment that can be done (with proper supervision) in an organic laboratory. Try dissolving benzoic acid crystals in room temperature water – you'll find that it is not soluble. As we will learn when we study acid-base chemistry in a later chapter, carboxylic acids such as benzoic acid are relatively weak acids, and thus exist mostly in the acidic (protonated) form when added to pure water.



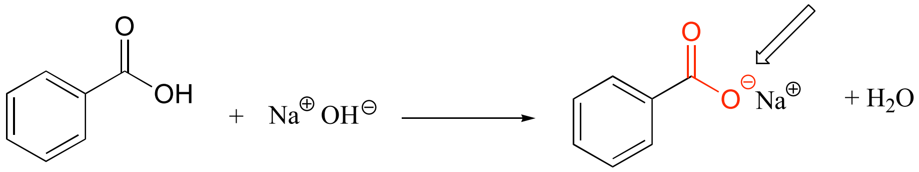
benzoic acid
insoluble in water



acetic acid
soluble in water

Acetic acid, however, is quite soluble. This is easy to explain using the small alcohol vs large alcohol argument: the hydrogen-bonding, hydrophilic effect of the carboxylic acid group is powerful enough to overcome the hydrophobic effect of a single methyl group on acetic acid, but not the larger hydrophobic effect of the 6-carbon benzene group on benzoic acid.

Now, try slowly adding some aqueous sodium hydroxide to the flask containing undissolved benzoic acid. As the solvent becomes more and more basic, the benzoic acid begins to dissolve, until it is completely in solution.



ionic - very hydrophilic!

benzoic acid
insoluble in water

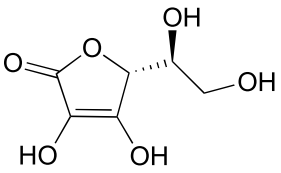
sodium benzoate
soluble in water

What is happening here is that the benzoic acid is being converted to its conjugate base, benzoate. The *neutral* carboxylic acid group was not hydrophilic enough to make up for the hydrophobic benzene ring, but the carboxylate group, with its *full negative charge*, is much more hydrophilic. Now, the balance is tipped in favor of water solubility, as the powerfully hydrophilic anion part of the molecule drags the hydrophobic part, kicking and screaming, (if a benzene ring can kick and scream) into solution. If you want to precipitate the benzoic acid back out of solution, you can simply add enough hydrochloric acid to neutralize the solution and reprotonate the carboxylate.

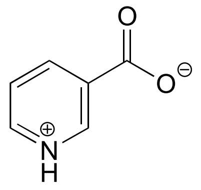
If you are taking a lab component of your organic chemistry course, you will probably do at least one experiment in which you will use this phenomenon to separate an organic acid like benzoic acid from a hydrocarbon compound like biphenyl.

Similar arguments can be made to rationalize the solubility of different organic compounds in nonpolar or slightly polar solvents. In general, the greater the content of charged and polar groups in a molecule, the less soluble it tends to be in solvents such as hexane. The ionic and very hydrophilic sodium chloride, for example, is not at all soluble in hexane solvent, while the hydrophobic biphenyl is very soluble in hexane.

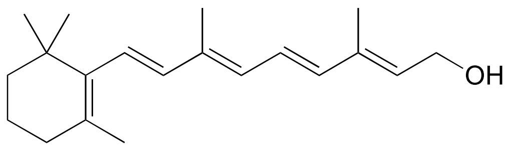
Exercise 2.12: Vitamins can be classified as water-soluble or fat-soluble (consider fat to be a very non-polar, hydrophobic 'solvent'. Decide on a classification for each of the vitamins shown below.



Vitamin C
(ascorbic acid)

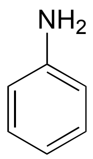


Vitamin B_3
(niacin)

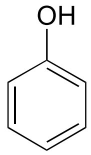


Vitamin A
(retinol)

Exercise 2.13: Both aniline and phenol are insoluble in pure water. Predict the solubility of these two compounds in 10% aqueous hydrochloric acid, and explain your reasoning. Hint – in this context, aniline is basic, phenol is not!



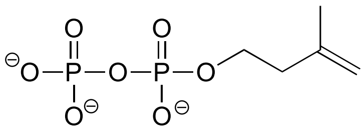
aniline



phenol

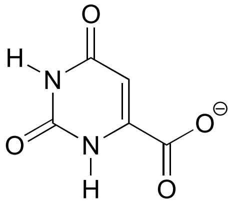
2.4B: Illustrations of solubility concepts: metabolic intermediates, lipid bilayer membranes, soaps and detergents

Because water is the biological solvent, most biological organic molecules, in order to maintain water-solubility, contain one or more charged functional groups. These are most often phosphate, ammonium or carboxylate, all of which are charged when dissolved in an aqueous solution buffered to pH 7.



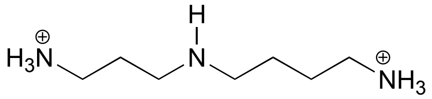
ispentenyl diphosphate

an early precuror to cholesterol



orotate

a precursor to the RNA base uridine

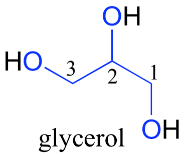


spermidine

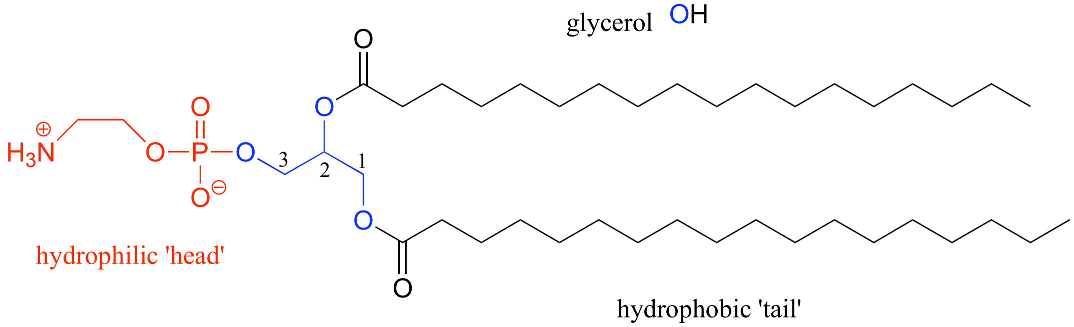
binds tightly to DNA in rapidly dividing cells

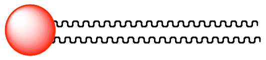
Sugars often lack charged groups, but as we discussed in our ‘thought experiment’ with glucose, they are quite water-soluble due to the presence of multiple hydroxyl groups.

Some biomolecules, in contrast, contain distinctly hydrophobic components. The ‘lipid bilayer’ membranes of cells and subcellular organelles serve to enclose volumes of water and myriad biomolecules in solution. The lipid (fat) molecules that make up membranes are **amphipathic**: they have a charged, hydrophilic ‘head’ and a hydrophobic hydrocarbon ‘tail’.



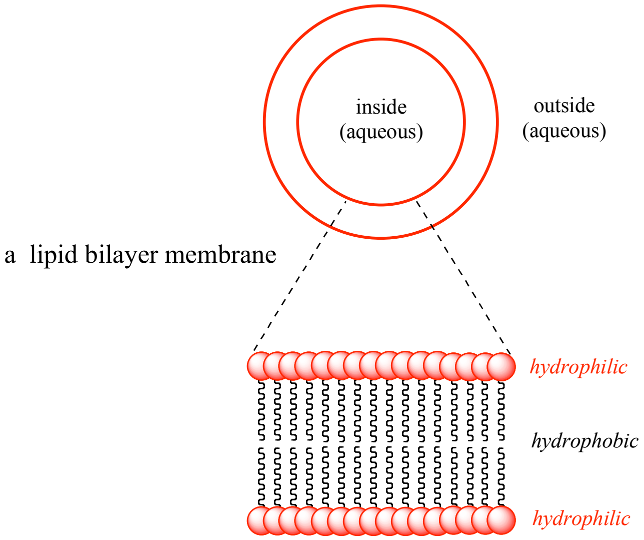
glycerol





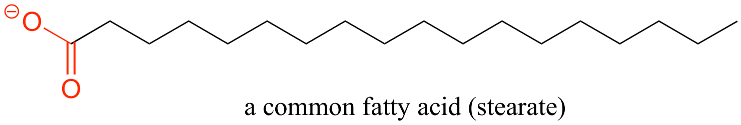
cartoon drawing of a membrane lipid

Notice that the entire molecule is built on a ‘backbone’ of glycerol, a simple 3-carbon molecule with three alcohol groups. In a biological membrane structure, lipid molecules are arranged in a spherical bilayer: hydrophobic tails point inward and bind together by van der Waals forces, while the hydrophilic head groups form the inner and outer surfaces in contact with water.



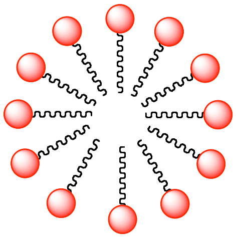
Because the interior of the bilayer is extremely hydrophobic, biomolecules (which as we know are generally charged species) are not able to diffuse through the membrane– they are simply not soluble in the hydrophobic interior. The transport of molecules across the membrane of a cell or organelle can therefore be accomplished in a controlled and specific manner by special transmembrane transport proteins, a fascinating topic that you will learn more about if you take a class in biochemistry.

A similar principle is the basis for the action of soaps and detergents. Soaps are composed of fatty acids, which are long (typically 18-carbon), hydrophobic hydrocarbon chains with a (charged) carboxylate group on one end,



a common fatty acid (stearate)

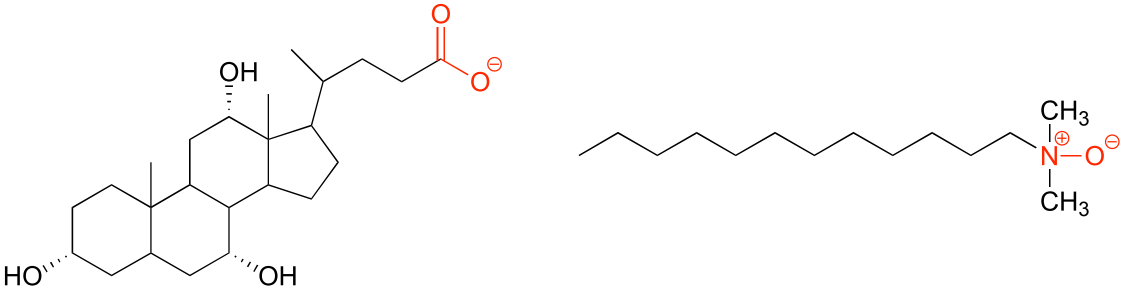
Fatty acids are derived from animal and vegetable fats and oils. In aqueous solution, the fatty acid molecules in soaps will spontaneously form **micelles**, a spherical structure that allows the hydrophobic tails to avoid contact with water and simultaneously form favorable van der Waals contacts.



micelle

Because the outside of the micelle is charged and hydrophilic, the structure as a whole is soluble in water. Micelles will form spontaneously around small particles of oil that normally would not dissolve in water (like that greasy spot on your shirt from the pepperoni slice that fell off your pizza), and will carry the particle away with it into solution. We will learn more about the chemistry of soap-making in a later chapter (section 12.4B).

Synthetic detergents are non-natural amphipathic molecules that work by the same principle as that described for soaps.

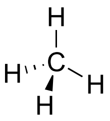


synthetic detergents

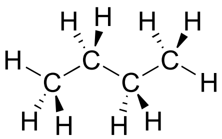
2.4C: Boiling points and melting points

The observable melting and boiling points of different organic molecules provides an additional illustration of the effects of noncovalent interactions. The overarching principle involved is simple: the stronger the noncovalent interactions between molecules, the more energy that is required, in the form of heat, to break them apart. Higher melting and boiling points signify stronger noncovalent intermolecular forces.

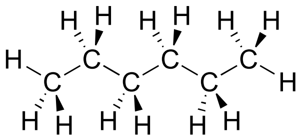
Consider the boiling points of increasingly larger hydrocarbons. More carbons means a greater surface area possible for hydrophobic interaction, and thus higher boiling points.



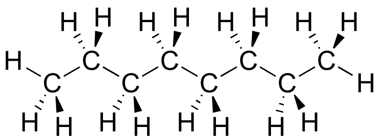
methane
bp = -164°C



butane
bp = 0°C

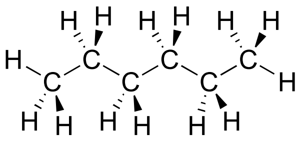


hexane
bp = 68°C

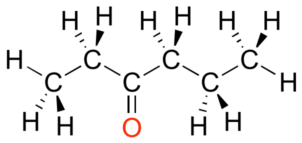


octane
bp = 128°C

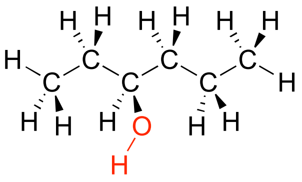
As you would expect, the strength of intermolecular hydrogen bonding and dipole-dipole interactions is reflected in higher boiling points. Just look at the trend for hexane (van der Waals interactions only), 3-hexanone (dipole-dipole interactions), and 3-hexanol (hydrogen bonding).



hexane
bp = 69^{o}C



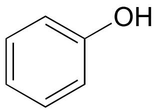
3-hexanone
bp = 123 ^{o}C

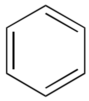


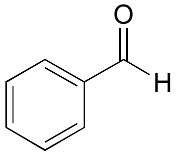
3-hexanol
bp =135^{o}C

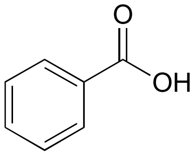
Of particular interest to biologists (and pretty much anything else that is alive in the universe) is the effect of hydrogen bonding in water. Because it is able to form tight networks of intermolecular hydrogen bonds, water remains in the liquid phase at temperatures up to 100 ^{O}C, (slightly lower at high altitude). The world would obviously be a very different place if water boiled at 30 ^{O}C.

Exercise 2.14: Based on their structures, rank phenol, benzene, benzaldehyde, and benzoic acid in terms of lowest to highest boiling point.

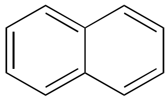




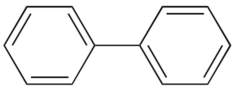




By thinking about noncovalent intermolecular interactions, we can also predict relative melting points. All of the same principles apply: stronger intermolecular interactions result in a higher melting point. Ionic compounds, as expected, usually have very high melting points due to the strength of ion-ion interactions (there are some ionic compounds, however, that are liquids at room temperature). The presence of polar and especially hydrogen-bonding groups on organic compounds generally leads to higher melting points. Molecular shape, and the ability of a molecule to pack tightly into a crystal lattice, has a very large effect on melting points. The flat shape of aromatic compounds such as napthalene and biphenyl allows them to stack together efficiently, and thus aromatics tend to have higher melting points compared to alkanes or alkenes with similar molecular weights.

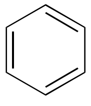


napthalene

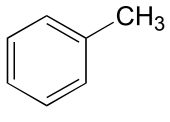


biphenyl

Comparing the melting points of benzene and toluene, you can see that the extra methyl group on toluene disrupts the molecule's ability to stack, thus decreasing the cumulative strength of intermolecular van der Waals forces.

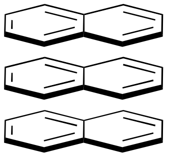


benzene
mp = 5.5 °C

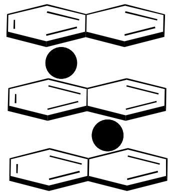


toluene
mp = -93 °C

If you are taking an organic lab course, you may have already learned that impurities in a crystalline substance will cause the observed melting point to be lower compared to a pure sample of the same substance. This is because impurities disrupt the ordered packing arrangement of the crystal, and make the cumulative intermolecular interactions weaker.



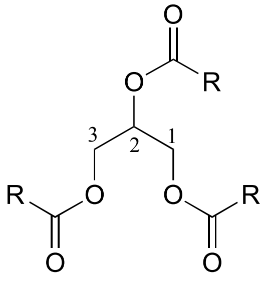
pure napthalene stacks efficiently:
mp = 80-82 °C



impurities disrupt the stacking, resulting in a lower mp

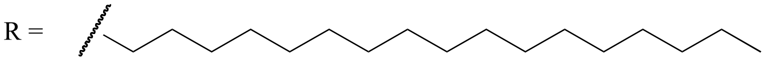
2.4D: The melting behavior of lipid structures

An interesting biological example of the relationship between molecular structure and melting point is provided by the observable physical difference between animal fats like butter or lard, which are solid at room temperature, and vegetable oils, which are liquid. Both solid fats and liquid oils are based on a 'triacylglycerol' structure, where three hydrophobic hydrocarbon chains of varying length are attached to a glycerol backbone through an ester functional group (compare this structure to that of the membrane lipids discussed in section 2.4B).

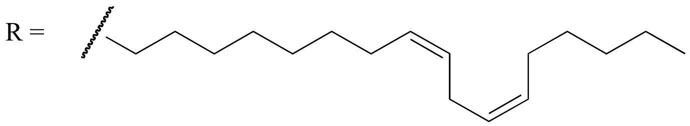


triacylglycerol

saturated fat:



unsaturated fat:

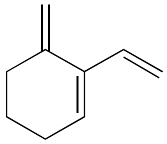


In vegetable oils, the hydrophobic chains are **unsaturated**, meaning that they contain one or more double bonds. Solid animal fat, in contrast, contains **saturated** hydrocarbon chains, with no double bonds. The double bonds in vegetable oils cause those hydrocarbon chains to be more rigid, and 'bent' at an angle (remember that rotation is restricted around double bonds), with the result that they don't pack together as closely, and thus can be broken apart (ie. melted) more readily.

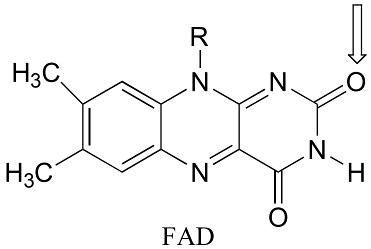
Chapter 2 problems

P2.1: Ozone, O_3, is an uncharged, non-cyclic molecule. Draw a Lewis structure for ozone. Are the two oxygen-oxygen bonds the likely to be the same length? What is the bond order? Explain.

P2.2: Redraw the structure below, showing the $2p_z$ orbitals that make up the conjugated π bond system (show the lowest energy molecular orbital).

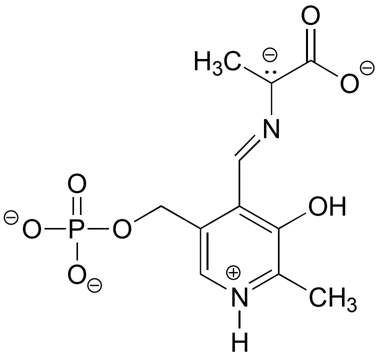


P2.3: Draw six different resonance contributors on the structure below (the flavin group of flavin adenine dinucleotide (FAD), a biochemical oxidation/reduction molecule) in which the oxygen indicated by an arrow bears a negative formal charge.



FAD

P2.4: The structure below shows an intermediate species in a reaction involving the amino acid alanine, attached to pyridoxal phosphate (vitamin B6). Draw a resonance contributor in which the only formal charges are on the oxygens.



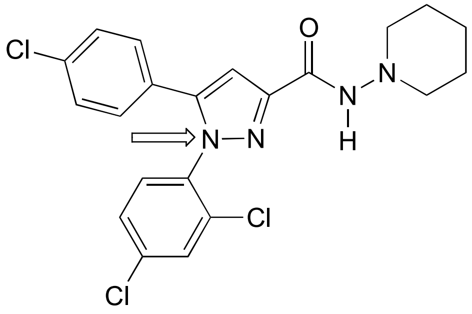
P2.5: In problem P1.8, you drew four different amides with molecular formula C_3H_7NO. One of these constitutional isomers is significantly less soluble in water than the other three. Which one, and why?

P2.6: Draw a MO diagram for a hypothetical He_2 molecule, and use this diagram to explain why there is no net lowering of energy when two helium atoms come together (and thus why He_2 is not observed). Use the MO diagram of H_2 in this chapter as a starting point.

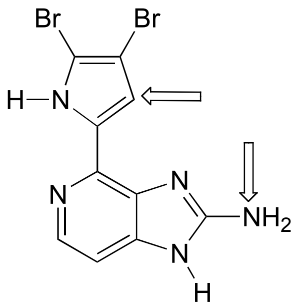
P2.7: Below is the structure of Rimonabant, a drug candidate which is being tested as a possible treatment for alcohol/tobacco dependence and obesity (see *Chemical and Engineering News*, October 15, 2006, p. 24). Draw resonance contributors in which

a) there is a separation of charge between the nitrogen indicated by an arrow and the oxygen.

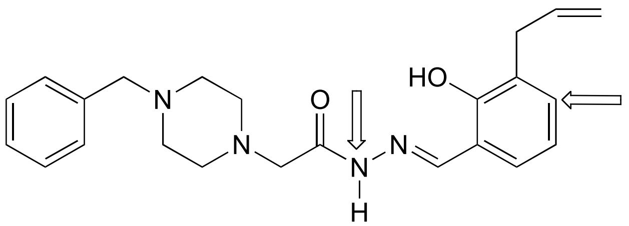
b) there is a separation of charge between a chlorine and one of the three nitrogens.



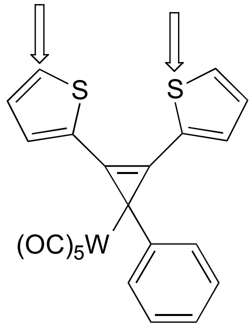
P2.8: For the molecules below, draw resonance contributors in which formal charges are placed on the atoms indicated by arrows. Use curved arrows to show how you are rearranging electrons between resonance contributors.



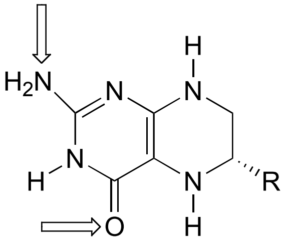
ageladine A, a potential cancer drug
Org. Lett. 2006, 8, 1443



PAC-1, a molecule that trigers cell death
Nature Chemical Biology 2006, 2, 543

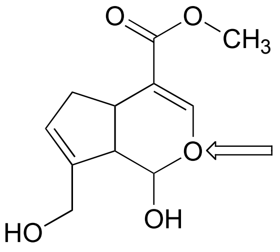


a new building block molecule for optoelectronic materials
Organometallics 2006, 25, 5176



tetrahydrofolate, a coenzyme

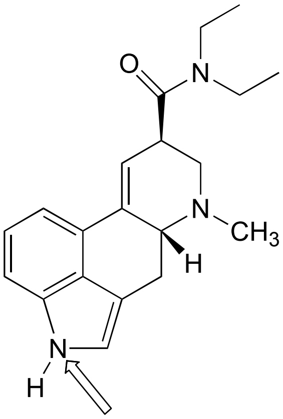
P2.9: Genepin was recently identified as the active compound in gardenia fruit extract, a traditional Chinese medicine for the treatment of diabetes (*Chemical and Engineering News* June 12, 2006, p. 34; *Cell Metab.* 2006, *3*, 417). Resonance contributors can be drawn in which the oxygen atom indicated by an arrow bears a positive formal charge. Indicate where the corresponding negative formal charge would be located in the most important of these contributors.



genipin

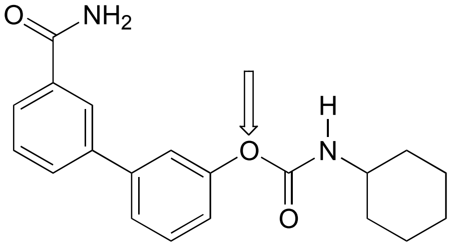
P2.10: Identify any *isolated* alkene groups in the PAC-1 structure in problem P2.8, and in the genipin structure in problem P2.9.

P2.11: The February 27, 2006 issue of *Chemical and Engineering News* contains an interesting article on the 100th birthday of Albert Hofmann, the inventor of the hallucinogen LSD. The structure of LSD is shown below. Several resonance contributors can be drawn in which the nitrogen atom indicated by an arrow bears a positive formal charge. Indicate atoms where a corresponding negative formal charge could be located in these contributors.



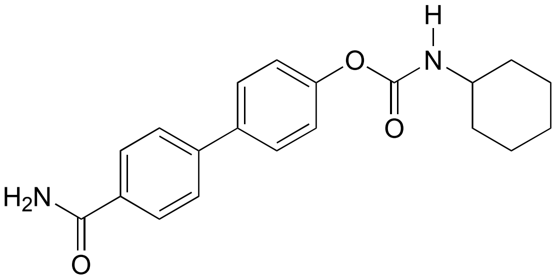
D-lysergic acid diethylamine (LSD)

P2.12: The human brain contains naturally occurring cannabinoid compounds which are related in structure to Δ^9-tetrahydrocannabinol, the active compound in cannabis. Cannabinoids are thought to exert an antidepressant effect. Because cannabis itself is not a realistic candidate for the treatment of depression due to its use as a recreational drug, researchers at the University of California, Irvine are studying synthetic compounds, such as the one shown below, which inhibit the degradation of natural cannabinoids in the brain. This compound has been shown to have antidepressant-like effects in rats and mice. (*Chemical and Engineering News*, December 19, 2005, p. 47; *Proc. Natl. Acad. Sci. USA* 2005, *102*, 18620.)



a) Several resonance contributors can be drawn in which the oxygen atom indicated by an arrow bears a positive formal charge. Indicate atoms where a corresponding negative formal charge could be located in these contributors.

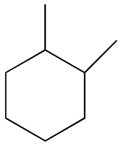
b) Answer the same question again, this time with the structural isomer shown below.

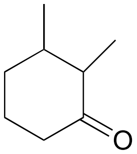


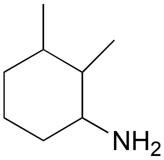
P2.13: Give the expected trend (lowest to highest) in boiling points for the following series of compounds :

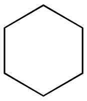
a) CH_3CHF_2 CH_2F_2 HF CH_3F

b)

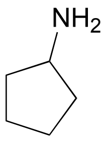


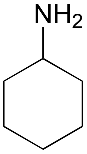


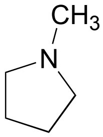




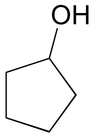
c)

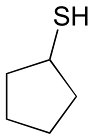


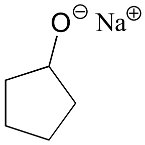


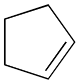


d)



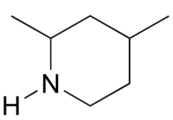




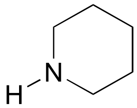


P2.14: For each pair of molecules below, choose the one that is more water-soluble, and explain your choice.

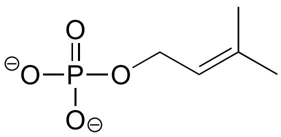
a)



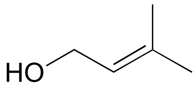
or



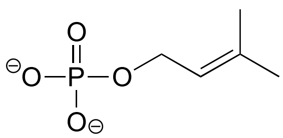
b)



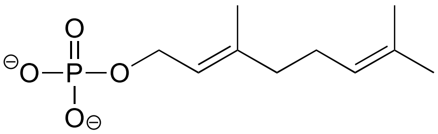
or



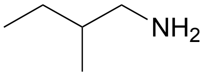
c)



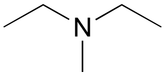
or



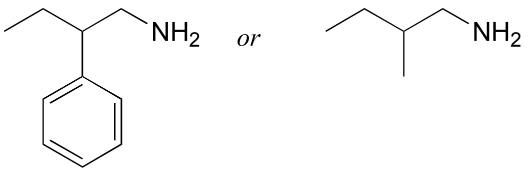
d)



or

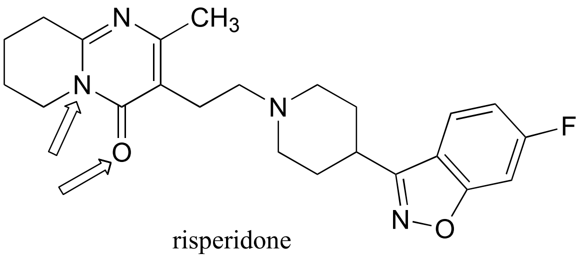


e)



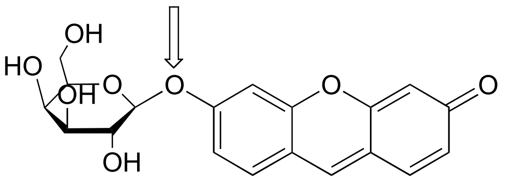
P2.15: Benzene and toluene (methylbenzene) are both commonly used organic solvents, although benzene is used less frequently in recent years due to its higher toxicity. Benzene has a boiling point of 80ºC and a melting point of 5.5 ºC. Toluene has a boiling point of 111 ºC and a melting point of -93 ºC. Use your knowledge of organic structure and noncovalent interactions to rationalize these trends in boiling and melting point.

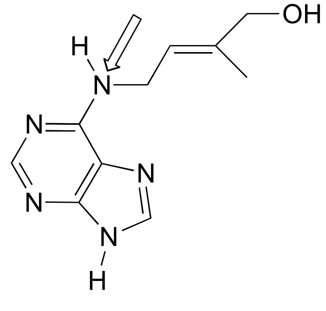
P2.16: Risperidone is a drug used for the treatment of schizophrenia and bipolar disorder. Draw two resonance contributors in which the nitrogen and oxygen atoms indicated both bear formal charges.



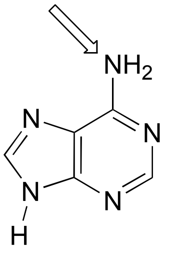
risperidone

P2.17: In each of the structures below, several resonance contributors can be drawn in which the atoms indicated by an arrow bear a positive formal charge. Circle all atoms which could bear the corresponding negative formal charges.



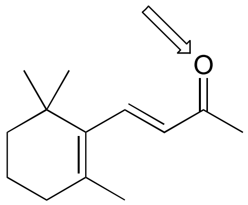


zeatin, a plant ripening hormone

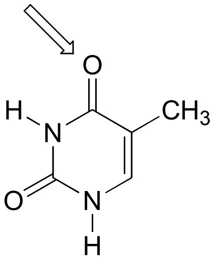


adenine

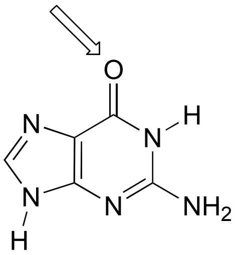
P2.18: In each of the structures below, several resonance contributors can be drawn in which the atoms indicated by an arrow bear a negative formal charge. Circle all atoms which could bear the corresponding positive formal charges.



B-ionone - tomato aroma

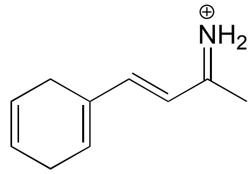


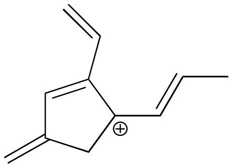
thymine



guanine

P2.19: In each of the structures below, circle all atoms to which the positive formal charge can be delocalized by resonance.



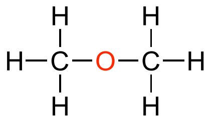


Chapter 3

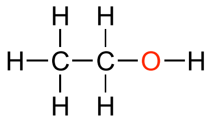
Conformation and stereochemistry

Introduction

In chapter 1, we learned that constitutional isomers are compounds with the same molecular formula and different bonding connectivity. Ethanol and dimethyl ether, for example, are constitutional isomers.



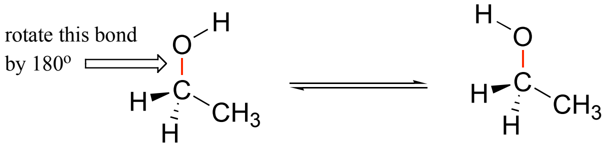
dimethyl ether
C_2H_6O



ethanol
C_2H_6O

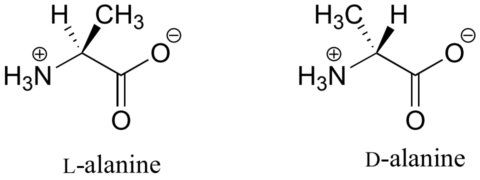
In this chapter, our goal is to learn about two more kinds of isomerism: conformational and stereochemical.

Conformational isomerism involves rotation about σ bonds, and does not involve any differences in the connectivity or geometry of bonding. Two or more structures that are categorized as conformational isomers, or **conformers**, are really just two of the exact same molecule that differ only in terms of the angle about one or more σ bonds. Here, for example, are two conformational isomers of ethanol in which the only difference is the angle of rotation about the C-O bond.



Although this example may seem trivial, we will see in this chapter and throughout the rest of our study of organic chemistry that understanding the energetic factors involved in the different possible conformations of a given molecule is critical to understanding the reactions that the molecule undergoes.

The bulk of this chapter is devoted to discussion of **stereoisomerism**, or **stereochemistry**. Two or more molecules are classified as stereoisomers if they have the same molecular formula and the same atom-to-atom connectivity, but have a *different arrangement of atoms in space*. Because of this differing bonding arrangement, two stereoisomers cannot be superimposed on one another – they are different molecules. The amino acid alanine, for example, has two stereoisomeric forms, which differ only in the bonding configuration at the central carbon:



L-alanine D-alanine

It is specifically the stereoisomer shown on the left-hand side of this figure, called L-alanine, which is found in virtually all proteins, in all forms of life on earth. The D-alanine isomer is quite rare.

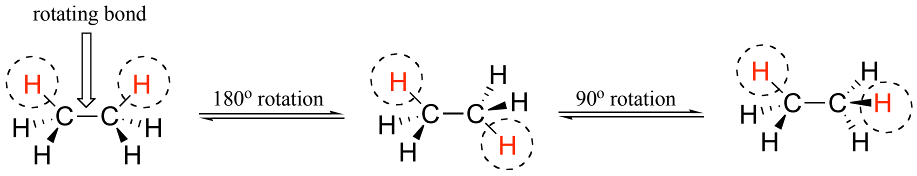
It is difficult to overstate the importance of stereoisomerism in biology and medicine: a difference in stereochemical configuration at a single carbon can make the difference between a new ‘wonder drug’ and an ineffective medication that causes dangerous side-effects.

Because this chapter deals with concepts that are inherently three-dimensional in nature, it will be very important for you to use a molecular modeling kit that is specifically intended for organic chemistry. Many of the concepts we will be exploring will be extremely confusing if you are limited to the two dimensions of this page – so now is the time to buy or borrow a good set of models if you don’t have one already!

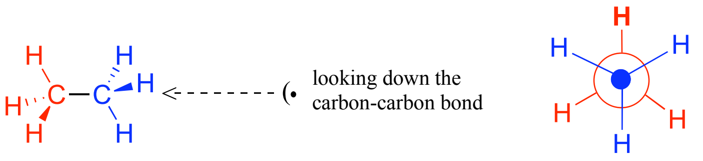
Section 3.1: Conformations of straight-chain organic molecules

3.1A: Conformations of ethane

Single bonds in organic molecules are free to rotate, with the result that organic molecules can adopt a huge number of possible conformations. Consider the very simple example of ethane. Although there are seven σ bonds in the ethane molecule, rotation about the six carbon-hydrogen bonds does not result in any change in the shape of the molecule because the hydrogen atoms are essentially spherical. Rotation about the carbon-carbon bond, however, results in many different possible molecular conformations.



In order to better visualize these different conformations, it is convenient to use a drawing convention called the **Newman projection**. In a Newman projection, we look lengthwise down a specific bond of interest – in this case, the carbon-carbon bond in ethane. We depict the 'front' atom as a dot, and the 'back' atom as a larger circle.

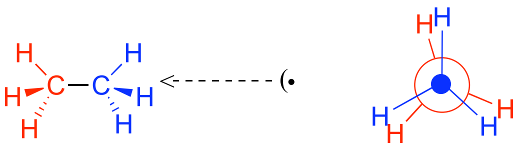


'staggered' conformation
(Newman projection)

The six carbon-hydrogen bonds are shown as solid lines protruding from the two carbons at 120° angles, which is what the actual tetrahedral geometry looks like when viewed from this perspective and flattened into two dimensions.

The lowest energy conformation of ethane, shown in the figure above, is called the 'staggered' conformation, in which all of the C-H bonds on the front carbon are positioned at dihedral angles of 60° relative to the C-H bonds on the back carbon. In this conformation, the distance between the bonds (and the electrons in them) is maximized.

If we now rotate the front CH_3 group 60° clockwise, the molecule is in the highest energy 'eclipsed' conformation, where the hydrogens on the front carbon are as close as possible to the hydrogens on the back carbon.



'eclipsed' conformation

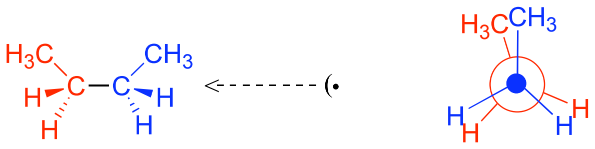
This is the highest energy conformation because of unfavorable interactions between the electrons in the front and back C-H bonds. The energy of the eclipsed conformation is approximately 3 kcal/mol higher than that of the staggered conformation.

Another 60° rotation returns the molecule to a second eclipsed conformation. This process can be continued all around the 360° circle, with three possible eclipsed conformations and three staggered conformations, in addition to an infinite number of variations in between.

The carbon-carbon bond is not *completely* free to rotate – there is indeed a small, 3 kcal/mol barrier to rotation that must be overcome for the bond to rotate from one staggered conformation to another. This rotational barrier is not high enough to prevent constant rotation except at extremely cold temperatures. However, at any given moment the molecule is more likely to be in a staggered conformation - one of the rotational 'energy valleys' - than in any other state.

3.1B: Conformations of butane

Now let us consider butane, a slightly larger molecule. There are now three rotating carbon-carbon bonds to consider, but we will focus on the middle bond between C_2 and C_3. Below are two representations of butane in a conformation which puts the two CH_3 groups (C_1 and C_4) in the eclipsed position.

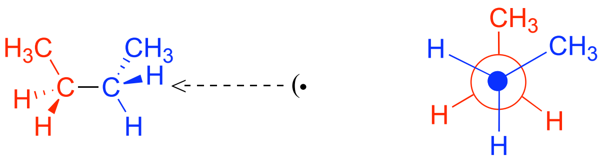


eclipsed (A)

This is the highest energy conformation for butane, due to what is called '**van der Waals repulsion**', or '**steric repulsion'**, between the two rather bulky methyl groups.

What is van der Waals repulsion? Didn't we just learn in Chapter 2 that the van der Waals force between two nonpolar groups is an *attractive* force? Consider this: you probably like to be near your friends, but no matter how close you are you probably don't want to share a one-room apartment with five of them. When the two methyl groups are brought too close together, the overall resulting noncovalent interaction is repulsive rather than attractive. The result is that their respective electron densities repel one another.

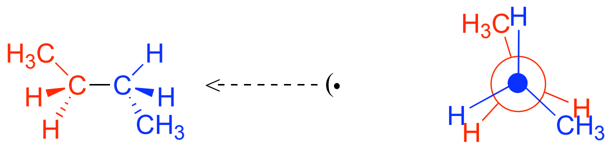
If we rotate the front, (blue) carbon by 60° clockwise, the butane molecule is now in a staggered conformation.



gauche

This is more specifically referred to as the '**gauche'** conformation of butane. Notice that although they are staggered, the two methyl groups are not as far apart as they could possibly be. There is still significant steric repulsion between the two bulky groups.

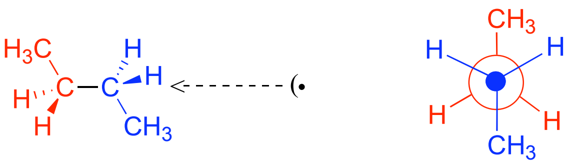
A further rotation of 60° gives us a second eclipsed conformation (B) in which both methyl groups are lined up with hydrogen atoms.



eclipsed (B)

Due to steric repulsion between methyl and hydrogen substituents, this eclipsed conformation B is higher in energy than the gauche conformation. However, because there is no methyl-to-methyl eclipsing, it is lower in energy than eclipsed conformation A.

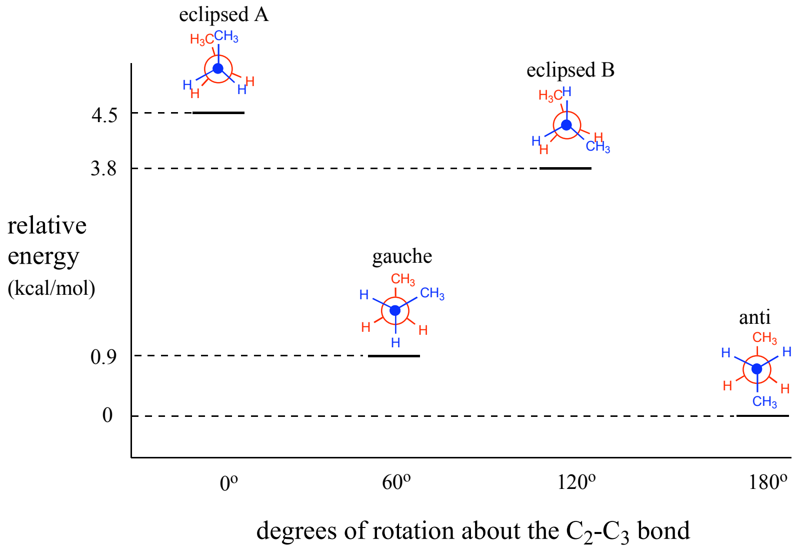
One more 60 rotation produces the ‘anti’ conformation, where the two methyl groups are positioned opposite each other and steric repulsion is minimized.



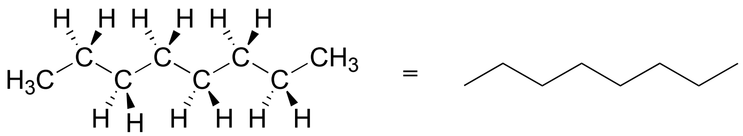
anti

This is the lowest energy conformation for butane.

The diagram below summarizes the relative energies for the various eclipsed, staggered, and gauche conformations.



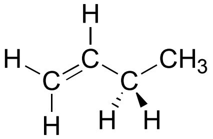
At room temperature, butane is most likely to be in the lowest-energy anti conformation at any given moment in time, although the energy barrier between the anti and eclipsed conformations is not high enough to prevent constant rotation except at very low temperatures. For this reason (and also simply for ease of drawing), it is conventional to draw straight-chain alkanes in a zigzag form, which implies anti conformation at all carbon-carbon bonds.



octane

Exercise 3.1: Draw Newman projections of the eclipsed and staggered conformations of propane.

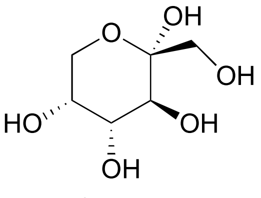
Exercise 3.2: Draw a Newman projection, looking down the C_2-C_3 bond, of 1-butene in the conformation shown below.



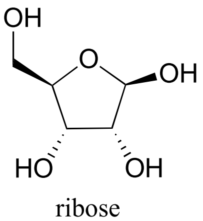
Section 3.2: Conformations of cyclic organic molecules

3.2A: Introduction to sugars and other cyclic molecules

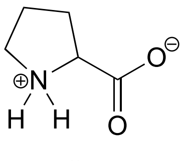
Browse through a biochemistry textbook and you will see any number of molecules with cyclic structures. Many of these cyclic structures are aromatic, and therefore planar. Many others, though, are non-aromatic, and it is these cyclic structures that are the topic of discussion in this section.



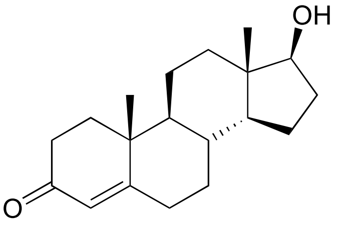
fructose



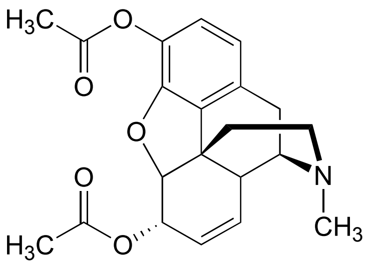
ribose



proline



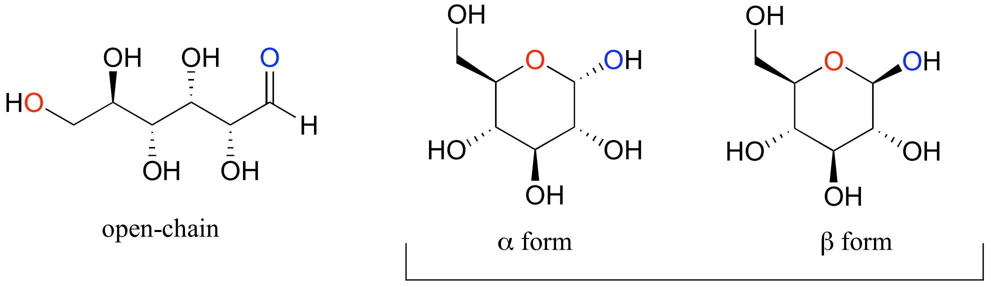
testosterone



heroin

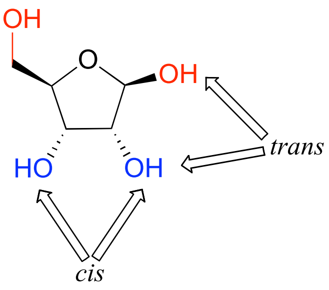
When discussing cyclic organic molecules, we will often use sugars as examples, because they are such important molecules in biological chemistry. It is important to understand that many sugars exist in aqueous solution as both straight-chain (also called 'open-chain') and cyclic forms. The open-chain form of glucose and two possible cyclic forms are shown below.

Different forms of glucose:



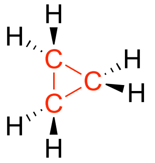
(Don't worry for now about how sugars convert between their open-chain and cyclic forms, or about the difference between the α and β forms of cyclic glucose - these and many other sugar-related topics will be covered later, mainly in chapter 11.)

One thing that you should notice in the cyclic structures shown above is that atoms or groups bonded to tetrahedral ring carbons are either pointing up (out of the plane of the page) or down (into the plane of the page), as indicated by the use of dashed or solid wedges. When two substituents on the same ring are both pointing in the same direction, they are said to be ***cis*** to each other. When they are pointed in opposite directions, they are said to be ***trans*** to each other.

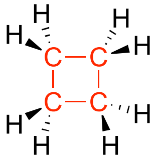


3.2B: Ring size

In most cases, ring structures in organic molecules are either five-membered or six-membered. Three-and four-membered rings are occasionally found in nature, but are significantly higher in energy. The relative instability of these smaller ring structures can be explained by a concept called **ring strain**. You can get a good idea of the source of ring strain if you try to build a model of cyclopropane or cyclobutane: in order to close the ring, the bonds between the carbons must be forced, or distorted, into angles quite a bit smaller than the tetrahedral angle of 109.5°.

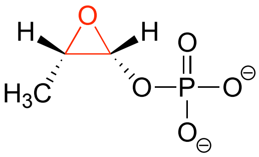


cyclopropane

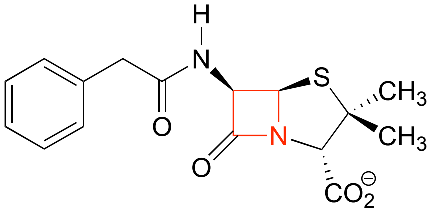


cyclobutane

If one of the carbon-carbon bonds is broken, the ring will 'spring' open, releasing energy as the bonds reassume their preferred tetrahedral geometry. The effectiveness of two antibiotic drugs, fosfomycin and penicillin, is due in large part to the high reactivity of the three- and four-membered rings in their structures.



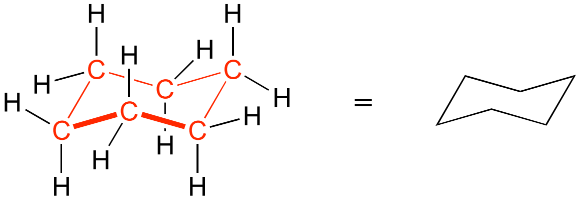
fosfomycin



penicillin G

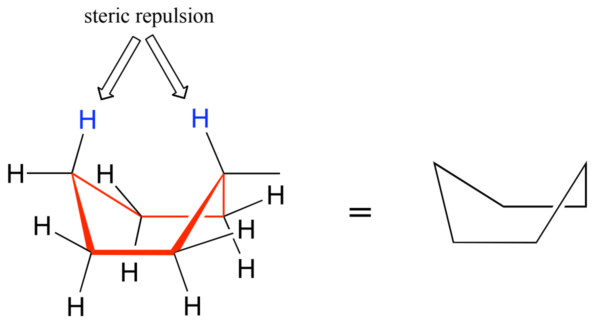
3.2C: Conformations of glucose and other six-membered ring structures

In the more open geometry of five- and six-membered cycloalkane structures, bonding angles are close to tetrahedral, and thus ring strain is not a factor – these rings are in fact very stable. However, the 'flat' drawings we have been seeing up to now do not accurately show the actual three-dimensional shape of a five- or six-membered ring. If cyclohexane were indeed flat, the bond angles would have to be distorted from 109.5° to 120°. If you build a model, though, you will find that if you rotate the carbon-carbon bonds so as to put the ring into a shape that resembles a reclining beach chair, all of the carbon-carbon bonds are able to assume tetrahedral bonding angles.



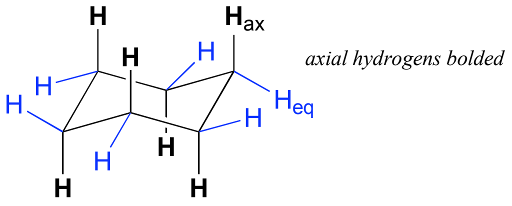
This ‘**chair’ conformation** is the lowest energy conformation for cyclohexane and other six-membered rings.

An alternate conformation for a six-membered ring is called the ‘**boat’**:

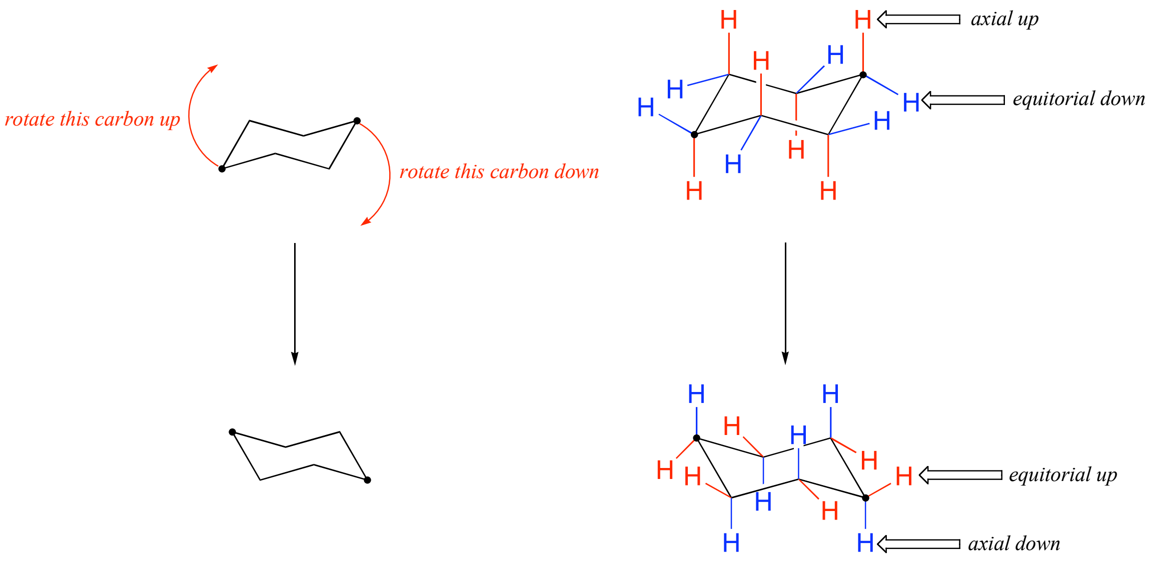


In the boat conformation, two of the substituents – those on the ‘bow’ and the ‘stern’ if you will – are brought close enough to each other to cause van der Waals repulsion. For this reason, the boat conformation is somewhat higher in energy than the chair conformation.

If you look carefully at your model of cyclohexane in the chair conformation, you will see that all twelve hydrogens are not equivalent in terms of their three-dimensional arrangement in space. Six hydrogens are **axial** – that is, they are pointing either straight up or straight down relative to the ring. The other six hydrogens are **equitorial**, meaning that they are pointing away from the perimeter of the ring, either slightly up or slightly down.

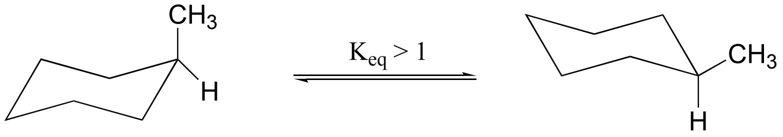


This is not the only possible chair conformation for cyclohexane. On your model, rotate one of the ‘up’ carbons down, and one of the ‘down' carbons up. You now have a new, alternate chair conformation – this process is called **ring inversion**.



What you should see is that, as a result of the ring inversion process, all of the axial and equitorial hydrogens have traded positions – axial hydrogens have become equitorial, and vice-versa. Notice, however, that the 'down' hydrogens are still pointing down, and the 'up' hydrogens are still pointing up regardless of whether they are axial or equitorial. At room temperature, cyclohexane is constantly inverting between two chair forms of equal energy – it is a rapid equilibrium situation. Thus, except at very low temperatures, we are not able to distinguish between axial and equitorial hydrogens, as they are constantly switching back and forth.

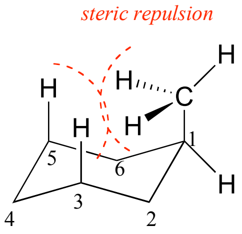
The picture is slightly different when we replace one of the hydrogen substituents with something bigger, such as a methyl group. Now, the two chair conformations are quite different: in one, the methyl group is equitorial and in the other it is axial.



methyl group axial

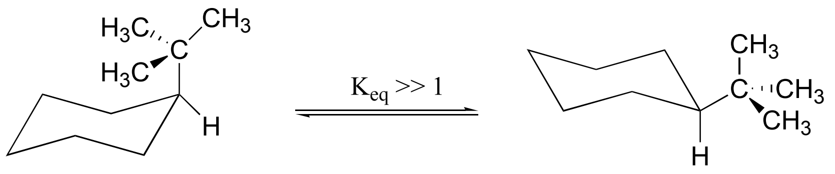
methyl group equitorial
(more stable)

When the methyl group is in the axial position, it is brought close enough to the hydrogens on carbons #3 and #5 to cause steric repulsion.



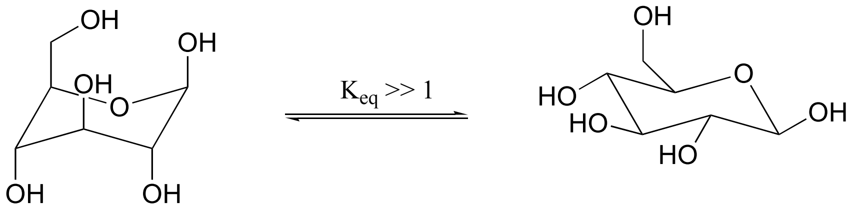
When in the equitorial position, the methyl group is pointing out *away* from the rest of the ring, eliminating this unfavorable interaction. As a consequence, the conformation in which the methyl group is in the equitorial position is more stable, by approximately 1.7 kcal/mol. At room temperature, methylcyclohexane exists as a rapid equilibrium between the two chair forms, but the equilibrium constant (K_{eq}) favors the conformation where the methyl group is equitorial.

The importance of the steric repulsion factor increases with the increasing size of a substituent. Tert-butyl cyclohexane, for example, is much more stable (by approximately 5.6 kcal/mol) in the chair conformation in which the bulky tert-butyl group is in the equitorial position.



As a general rule, *the most stable chair conformation of a six-membered ring will be that in which the bulkier groups are in the equitorial position.*

Previously, we have seen the six ring atoms of cyclic glucose drawn in two dimensions. A more accurate depiction shows that the molecule adopts, as expected, a chair conformation.

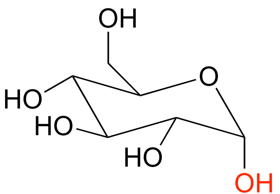


all axial

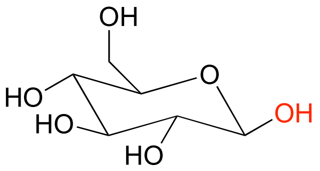
all equitorial

The conformation in which all substituents are equatorial is lower in energy.

We will learn later in chapter 11 that glucose can also exist in an alternate cyclical form. The two isomeric forms are referred to by the Greek letters α and β.



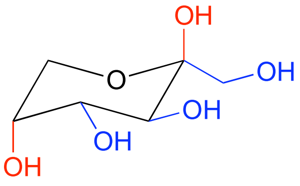
α-glucose



β-glucose

We have not learned about stereoisomerism quite yet, but you can still recognize that the bonding configuration on one carbon is different. On the α isomer, one of the hydroxyls is axial – this isomer is not able to adopt a chair conformation in which all non-hydrogen substituents are equitorial. The lower energy conformation is the one in which four of the five substituents are equitorial, but the presence of the one axial hydroxyl group means that the α isomer is, overall, less stable than the β isomer.

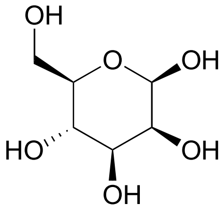
The most abundant form of fructose in aqueous solution is also a six-membered ring.



fructose
(major form)

The lower energy chair conformation is the one with three of the five substituents (including the bulky –CH_2OH group) in the equitorial position.

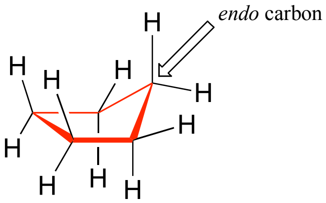
Exercise 3.3: Draw the two chair conformations of the sugar called mannose, being sure to clearly show each non-hydrogen substituent as axial or equitorial. Predict which conformation is likely to be more stable, and explain why.



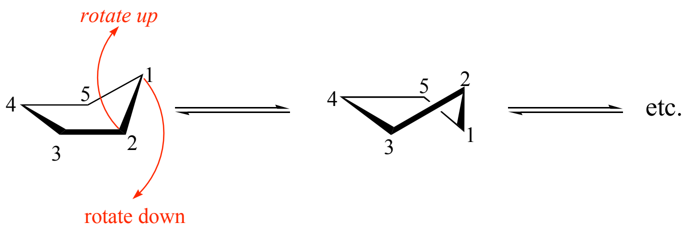
mannose

3.2D: Conformations of pentose and other five-membered ring structures

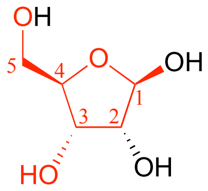
The lowest energy conformation of cyclopentane and other five-membered rings is known as the 'envelope', with four of the ring atoms in the same plane and one out of plane (notice that this shape resembles an envelope with the flap open). The out-of-plane carbon is said to be in the *endo* position ('*endo'* means 'inside').



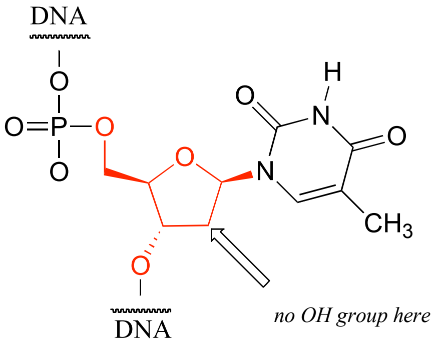
At room temperature, cyclopentane undergoes a rapid '**pseudorotation'** process in which each of the five carbons takes turns being in the *endo* position.



One of the most important five-membered rings in nature is a sugar called ribose – DNA and RNA are both constructed upon 'backbones' that are derived from ribose. Pictured below is one thymidine (T) deoxy-nucleotide from a stretch of DNA:

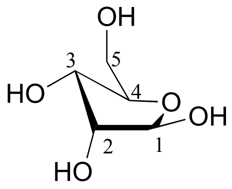


ribose



no OH group here in DNA

The lowest-energy conformations for ribose are envelope forms in which either C_3 or C_2 are *endo*, on the same side as the C_5 substituent.



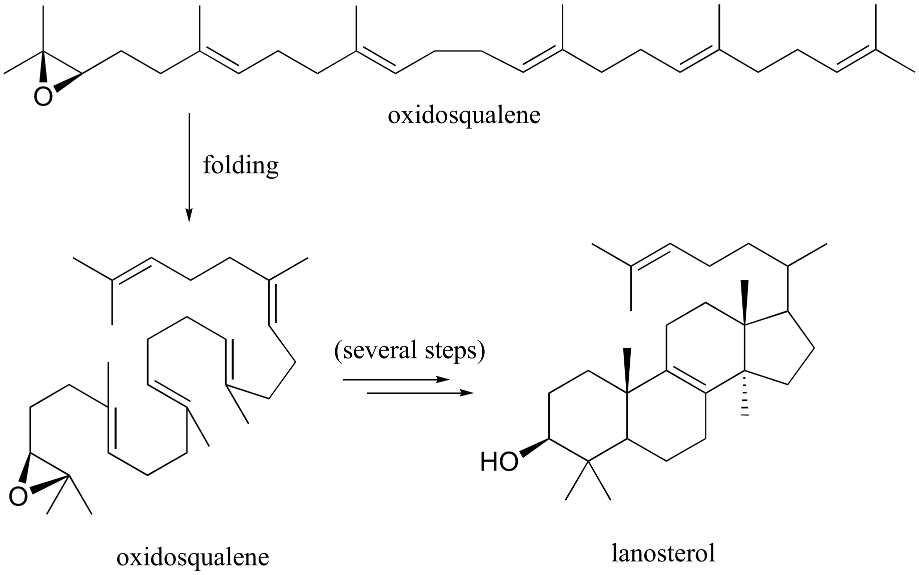
C_3-endo conformation (as in RNA)

This has very important implications for oligonucleotide structure – in DNA, it is C_2 that is in the *endo* position, while in RNA it is C_3.

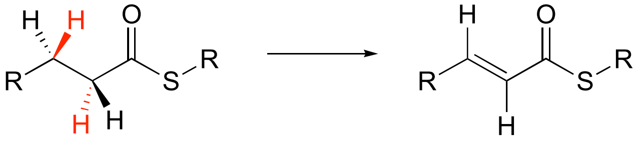
Exercise 3.4: Draw the C_2-*endo* form of ribose.

3.2E: The importance of conformation in organic reactivity

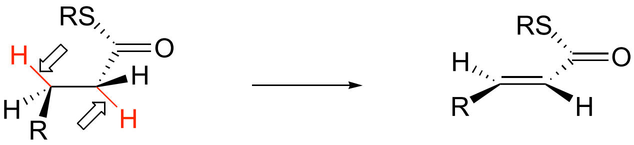
As we continue to study the structure and reactivity of organic molecules, and biomolecules in particular, we will see many more examples in which questions of bond rotations and molecular conformation are critical. Many reactions will not proceed unless the starting compound is in a very specific starting conformation. Oxidosqualene, in order to undergo a complex cyclization reaction to form lanosterol, must first adopt - with the help of an enzyme - a very specific conformation (this reaction, which is part of the biosynthetic pathway of cholesterol, will be discussed in section 15.7B).



The reaction shown below is part of the process by which our bodies burn fats for energy (we'll discuss this reaction in detail in section 16.5).



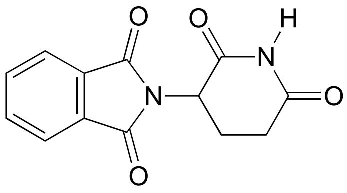
The fat molecule is held by an enzyme in a very specific conformation. The two C-H bonds that break must lie in the same plane in order for the reaction to occur.



indicated bonds must be in the same plane

Section 3.3: Stereoisomerism – chirality, stereocenters, enantiomers

In the late 1950's and early 1960's, a drug called thalidomide was used in many European countries to alleviate morning sickness in pregnant women.

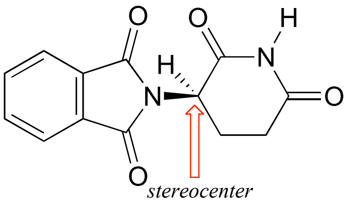


thalidomide

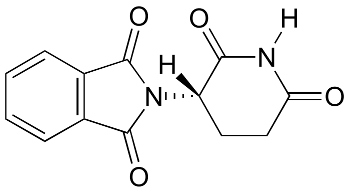
It was not long, however, before doctors realized that something had gone horribly wrong: many babies born to women who had taken thalidomide during pregnancy suffered from severe birth defects.

Researchers later realized the that problem lay in the fact that thalidomide was being provided as a mixture of two different isomeric forms.

effective isomer



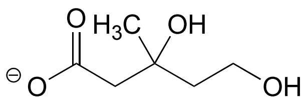
mutagenic isomer



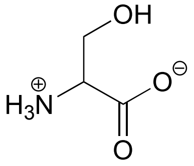
One of the isomers is an effective medication, the other caused the side effects. Both isomeric forms have the same molecular formula and the same atom-to-atom connectivity, so they are not constitutional isomers. Where they differ is in the arrangement in three-dimensional space about one tetrahedral, sp^3-hybridized carbon. These two forms of thalidomide are **stereoisomers**.

Note that the carbon in question has *four different substituents* (two of these just happen to be connected by a ring structure). Tetrahedral carbons with four different substituent groups are called **stereocenters**.

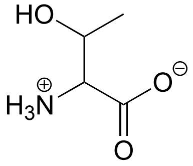
Exercise 3.5: Locate all of the carbon stereocenters in the molecules below.



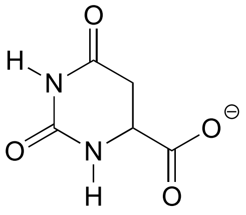
mevalonate



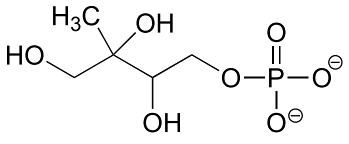
serine



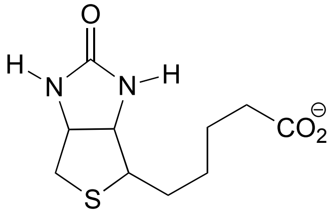
threonine



dihydroorotate

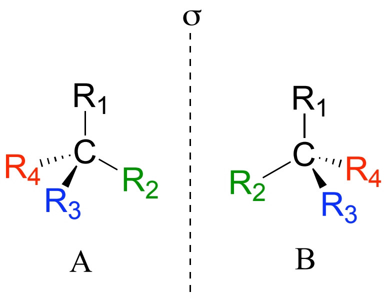


2-methylerythritol-4-phosphate

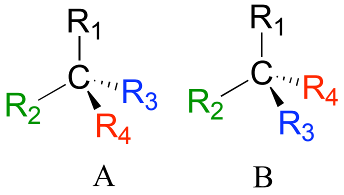


biotin

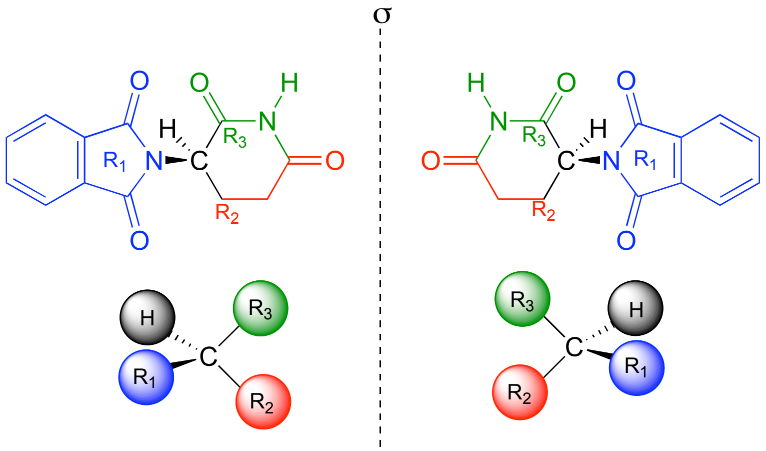
Looking at the structures of what we are referring to as the two isomers of thalidomide, you may not be entirely convinced that they are actually two different molecules. In order to convince ourselves that they are indeed different, let's create a generalized picture of a tetrahedral carbon stereocenter, with the four substituents designated R_1-R_4. The two stereoisomers of our simplified model look like this:



If you look carefully at the figure above, you will notice that molecule A and molecule B are mirror images of each other (the line labeled 'σ' represents a mirror plane). Furthermore, *they are not superimposable*: if we pick up molecule A, flip it around, and place it next to molecule B, we see that the two structures cannot be superimposed on each other. They are different molecules!



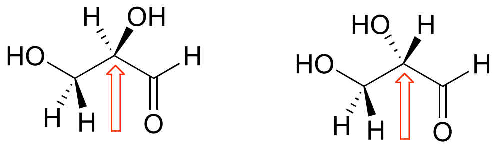
If you make models of the two stereoisomers of thalidomide and do the same thing, you will see that they too are mirror images, and cannot be superimposed (it well help to look at a color version of the figure below).

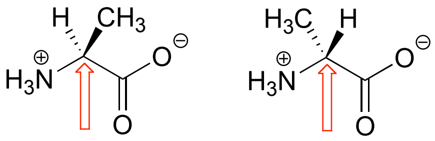


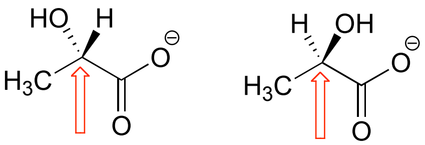
Thalidomide is a **chiral** molecule. Something is considered to be chiral if it cannot be superimposed on its own mirror image – in other words, if it is **asymmetric** (lacking in symmetry). The term 'chiral' is derived from the Greek word for 'handedness' – ie. right-handedness or left-handedness. Your hands are chiral: your right hand is a mirror image of your left hand, but if you place one hand on top of the other, both palms down, you see that they are not superimposable.

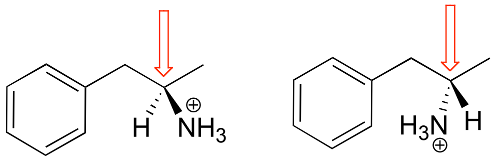
A pair of stereoisomers that are non-superimposable mirror images of one another are considered to have a specific type of stereoisomeric relationship – they are a pair of **enantiomers**. Thalidomide exists as a pair of enantiomers. On the macro level, your left and right hands are also a pair of enantiomers.

Here are some more examples of chiral molecules that exist as pairs of enantiomers. In each of these examples, there is a single stereocenter, indicated with an arrow. (Many molecules have more than one stereocenter, but we will get to that that a little later!)

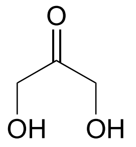

glyceraldehyde


alanine

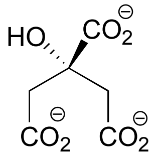

lactate


amphetamine

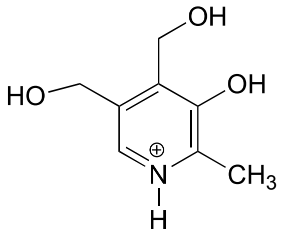
Here are some examples of molecules that are **achiral** (*not* chiral). Notice that none of these molecules has a stereocenter.



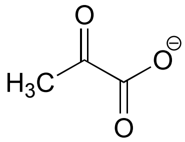
dihydroxyacetone



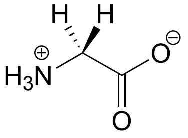
citrate



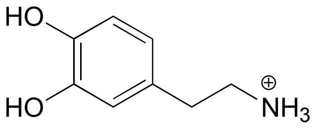
pyridoxine
(vitamin B_6)



pyruvate



glycine
(amino acid)



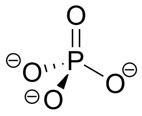
dopamine
(neurotransmitter)

It is difficult to illustrate on the two dimensional page, but you will see if you build models of these achiral molecules that, in each case, there is at least one **plane of symmetry**, where one side of the plane is the mirror image of the other. Chirality is tied conceptually to the idea of asymmetry, and *any molecule that has a plane of symmetry cannot be chiral*. When looking for a plane of symmetry, however, we must consider all possible conformations that a molecule could adopt. Even a very simple molecule like ethane, for example, is asymmetric in many of its countless potential conformations – but it has obvious symmetry in both the eclipsed and staggered conformations, and for this reason it is achiral.

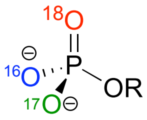
Looking for planes of symmetry in a molecule is useful, but often difficult in practice. In most cases, the easiest way to decide whether a molecule is chiral or achiral is to look for one or more stereocenters - with a few rare exceptions (see section 3.7B), the general rule is that molecules with at least one stereocenter are chiral, and molecules with no stereocenters are achiral. Carbon stereocenters are also referred to quite frequently as **chiral carbons**.

When evaluating a molecule for chirality, it is important to recognize that the question of whether or not the dashed/solid wedge drawing convention is used is irrelevant. Chiral molecules are sometimes drawn without using wedges (although obviously this means that stereochemical information is being omitted). Conversely, wedges may be used on carbons that are not stereocenters – look, for example, at the drawings of glycine and citrate in the figure above. Just because you see dashed and solid wedges in a structure, do not automatically assume that you are looking at a stereocenter.

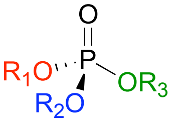
Other elements in addition to carbon can be stereocenters. The phosphorus center of phosphate ion and organic phosphate esters, for example, is tetrahedral, and thus is potentially a stereocenter.



phosphate
(achiral)



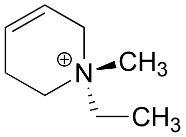
labeled phosphate ester
(chiral)



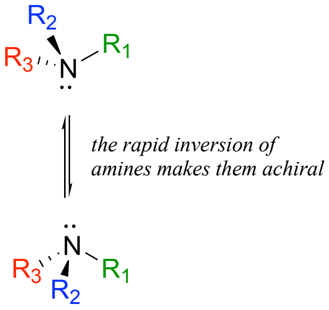
a chiral phosphate triester

We will see in chapter 10 how researchers, in order to investigate the stereochemistry of reactions at the phosphate center, incorporated sulfur and/or ^{17}O and ^{18}O isotopes of oxygen (the 'normal' isotope is ^{16}O) to create chiral phosphate groups. Phosphate triesters are chiral if the three substituent groups are different.

Asymmetric quaternary ammonium groups are also chiral. Amines, however, are not chiral, because they rapidly invert, or turn 'inside out', at room temperature.



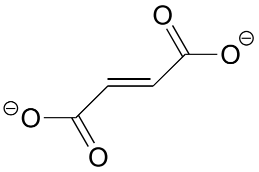
a chiral quaternary ammonium



the rapid inversion of amines makes them achiral

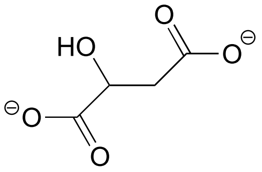
Exercise 3.6: Label the molecules below as chiral or achiral, and locate all stereocenters.

a)



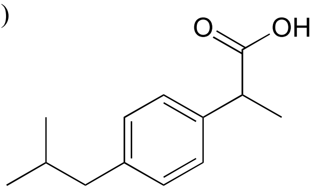
fumarate
(a citric acid cycle intermediate)

b)



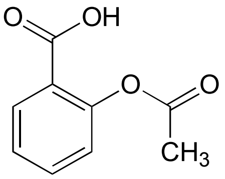
malate
(a citric acid cycle intermediate)

c)



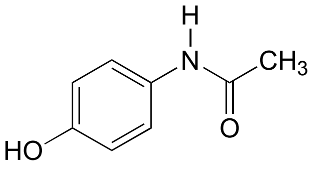
ibuprofin

d)



acetylsalicylic acid
(aspirin)

e)

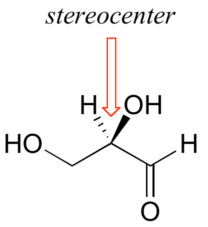


acetaminophen
(active ingredient in Tylenol)

Section 3.4: Defining stereochemical configuration - the Cahn-Ingold-Prelog system

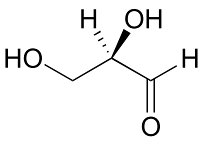
Chemists need to have a convenient way to distinguish between stereoisomers. The **Cahn-Ingold-Prelog system** is a set of rules that allows us to unambiguously define the stereochemical configuration of any stereocenter, using the designations '**R**' (from the Latin *rectus*, meaning right-handed) or '**S**' (from the Latin *sinister*, meaning left-handed).

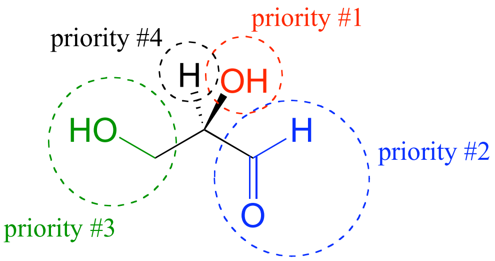
The rules for this system of stereochemical nomenclature are, on the surface, fairly simple. Because thalidomide is a somewhat complicated molecule, we'll use the simple 3-carbon sugar glyceraldehyde as our first example. You will want to make a model of the stereoisomer of glyceraldehyde shown below (be sure that you are making the correct enantiomer!).

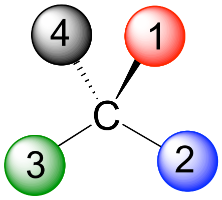


glyceraldehyde
(one enantiomer)

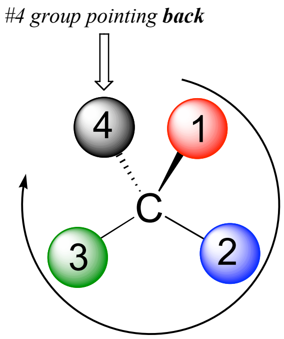
The first thing that we must do is to assign a **priority** to each of the four substituents bound to the **chiral carbon**. In this nomenclature system, the priorities are based on atomic number, with higher atomic numbers having a higher priority. We first look at the atoms that are directly bonded to the chiral carbon: these are H, O (in the hydroxyl), C (in the aldehyde), and C (in the CH_2OH group).



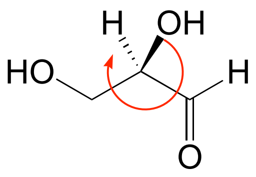




Two priorities are easy: hydrogen, with an atomic number of 1, is the lowest (#4) priority, and the hydroxyl oxygen, with atomic number 8, is priority #1. Carbon has an atomic number of 6. Which of the two 'C' groups is priority #2, the aldehyde or the CH_2OH? To determine this, we move one more bond away from the stereocenter: for the aldehyde we have a *double* bond to an oxygen, while on the CH_2OH group we have a *single* bond to an oxygen. If the atom is the same, double bonds have a higher priority than single bonds. Therefore, the aldehyde group is assigned #2 priority and the CH_2OH group the #3 priority. With our priorities assigned, we next make sure that the #4 priority group (the hydrogen) is pointed back away from ourselves, into the plane of the page (it is already).

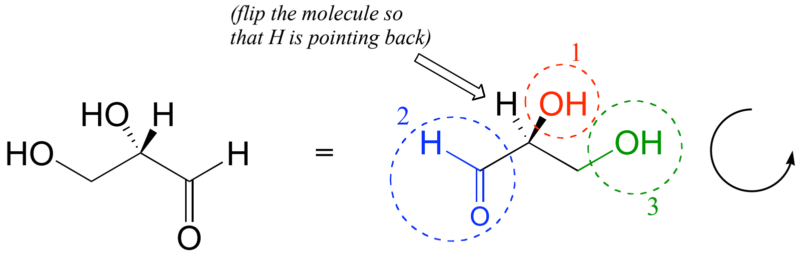


clockwise = R configuration



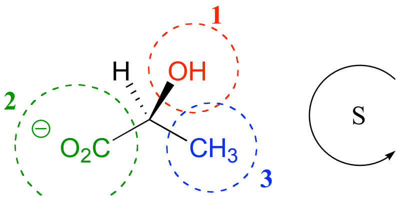
(R)-glyceraldehyde

Then, we trace a circle defined by the #1, #2, and #3 priority groups, in increasing order. For our glyceraldehyde example, this circle is clockwise, which tells us that this carbon has the 'R' configuration, and that this molecule is (R)-glyceraldehyde. For (S)-glyceraldehyde, the circle described by the #1, #2, and #3 priority groups is counter-clockwise (but first, we must flip the molecule over so that the H is pointing into the plane of the page).



(S)-glyceraldehyde

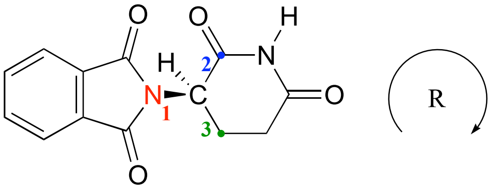
We can follow a similar procedure for a molecule such as lactate. Clearly, H is the #4 substituent and OH is #1. Owing to its three bonds to oxygen, the CO_2 group takes priority #2.



(S)-lactate

The structure shown is the S enantiomer of lactate.

Now, let's try to assign a stereochemical configuration to the enantiomer of thalidomide that is the effective medication. The #4 priority is again hydrogen (in fact, the #4 group will be hydrogen for most of the stereocenters we encounter). The nitrogen group is #1, the carbonyl side of the ring is #2, and the –CH_2 side of the ring is #3.



(R)-thalidomide

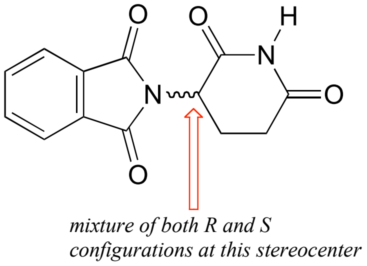
The hydrogen is shown pointing away from us, and the prioritized substituents trace a clockwise circle: this is the R enantiomer of thalidomide.

Exercise 3.7: Label stereochemistry for the figures of glyceraldehyde, alanine, lactate and amphetamine shown earlier in the chapter (section 3.3).

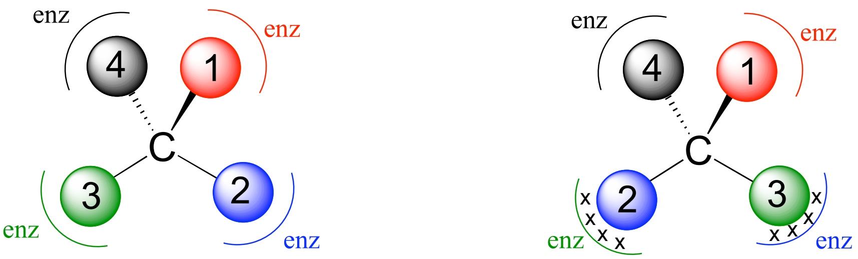
Exercise 3.8: Using solid or dashed wedges to show stereochemistry, draw the (R)-enantiomers of malate and ibuprofin (structures in exercise 3.6), and mevalonate, 2-methylerithritol-4-phosphate, and dihydroorotate (structures in section 3.3).

Section 3.5: Interactions between chiral molecules and proteins

The thalidomide that was used in the 1960s to treat depression and morning sickness was sold as a 50:50 mixture of both the R and the S enantiomer – this is referred to as a **racemic mixture**. A **squiggly bond** in a chemical structure indicates a racemic mixture - thus racemic (R/S) thalidomide would be drawn as:


O
O
H
N
N
O
O
mixture of both R and S
configurations at this stereocenter

The problem with racemic thalidomide, as we know, was that only the R enantiomer was an effective medicine, while the S enantiomer caused mutations in the developing fetus. How does such a seemingly trivial structural variation lead to such a dramatic (and in this case, tragic) difference in biological activity? Virtually all drugs work by interacting in some way with important proteins in our cells: they may bind to pain receptor proteins to block the transmission of pain signals, for instance, or clog up the active site of an enzyme that is involved in the synthesis of cholesterol. Proteins are chiral molecules, and are very sensitive to stereochemistry: just as a right-handed glove won't fit on your left hand, a protein that is able to bind tightly to (R)-thalidomide may not bind well at all to (S)-thalidomide (it will help to view a color version of the figure below).

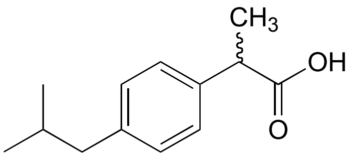


a protein that binds specifically to a chiral molecule . . .

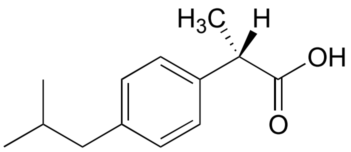
may not bind well at all to a different stereoisomer!

Instead, it seems that (S)-thalidomide interacts somehow with a protein involved in the development of a growing fetus, eventually causing the observed birth defects.

The over-the-counter painkiller ibuprofen is currently sold as a racemic mixture, but only the S enantiomer is effective.



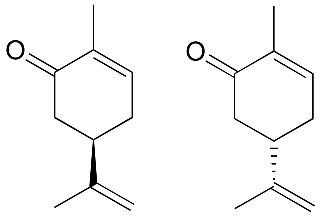
racemic ibuprofin



(S)-ibuprofin
(active enantiomer)

Fortunately, the R enantiomer does not produce any dangerous side effects, although its presence does seem to increase the amount of time that it takes for (S)-ibuprofen to take effect.

You can, with the assistance your instructor, directly experience the biological importance of stereoisomerism. Carvone is a chiral, plant-derived molecule that smells like spearmint in the R form and caraway (a spice) in the S form.



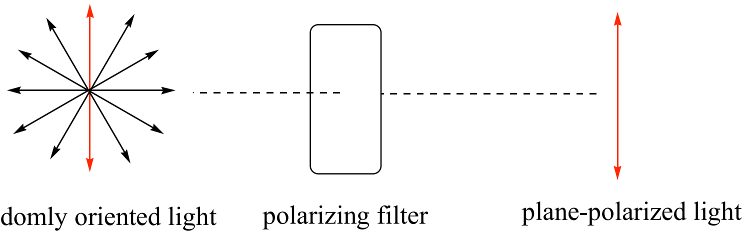
(R)-carvone (S)-carvone

The two enantiomers interact differently with smell receptor proteins in your nose, generating the transmission of different chemical signals to the olfactory center of your brain.

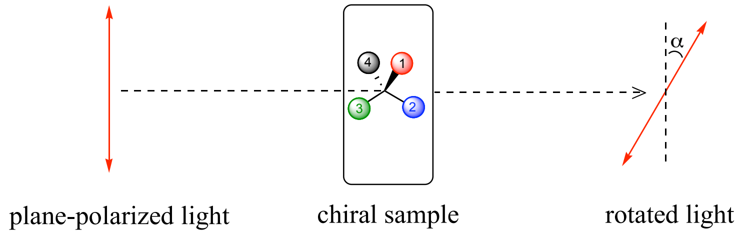
Section 3.6: Optical activity

While your nose, with the help of its chiral smell receptor proteins, can tell the difference between (R)-carvone and (S)-carvone, someone with a bad cold would have a difficult time distinguishing between the two enantiomers using most standard organic laboratory techniques. *Enantiomers have identical physical properties*: the same melting point, boiling point, solubility behavior, chromotographic mobility, index of refractivity, etc. Without the help of a specific chiral agent (e.g. smell receptor proteins) one of the only ways in which a pair of enantiomers can be conveniently distinguished in the laboratory is by a method known as **polarimetry**.

You may know from studying physics that light is characterized by electric and magnetic fields that oscillate in planes that are perpendicular to one another. In 'normal', unpolarized light, these oscillations are randomly oriented in all planes. Light that has been passed through a polarizing filter oscillates in only one direction, and is called **plane polarized light**.



Chiral molecules have a property known as **optical activity**: what this means is that a beam of plane-polarized light, when passed through a sample of a chiral substance, will interact with the substance in such a way that the angle of the plane of oscillation will rotate.



Pure enantiomers will always rotate plane-polarized light to an equal but opposite degree. If a stereochemically pure sample of (S)-carvone, for example, rotates plane-polarized light by +10° (clockwise), then a sample of (R)-carvone (in the exact same concentration and under the same experimental conditions) will rotate light by –10° (counterclockwise).

The magnitude of the observed optical activity is dependent on temperature, the wavelength of light used, solvent, concentration of the chiral sample, and the path length of the sample tube. (Path length is the length, expressed in decimeters (1 dm – 10 cm), that the plane-polarized light travels through the chiral sample). Typically, optical activity measurements are made in a 1 decameter (10 cm) path-length sample tube at 25 °C, using as a light source the so-called "D-line" from a sodium lamp, which has a wavelength of 589 nm and is yellow in color. The **specific rotation** [α] of a compound at 25° is expressed by:

$$[\alpha]_D^{25} = \frac{\alpha}{lc}$$

where α is the observed rotation, *l* is path length in dm, and *c* is the concentration of the sample in grams per 100 mL. Because every chiral molecule has a characteristic specific rotation, polarimetry measurements can be used to determine the concentration of a pure sample of a chiral compound in solution, assuming that its specific rotation is already known. This is a common method for determining the concentration of sugar solutions.

Because R and S enantiomers have equal but opposite optical activity, it naturally follows that a 50:50 racemic mixture of two enantiomers will have no observable optical activity. If we know the specific rotation for a chiral molecule, however, we can easily calculate the ratio of enantiomers present in a mixture of two enantiomers, based on its measured optical activity. When a mixture contains more of one enantiomer than the other, chemists often use the concept of **enantiomeric excess (ee)** to quantify the difference. Enantiomeric excess can be expressed as:

$$\text{ee} = \frac{(\text{\% more abundant enantiomer - 50) x 100}}{50}$$

For example, a mixture containing 60% *R* enantiomer (and 40% *S* enantiomer) has a 20% enantiomeric excess of R: ((60-50) x 100) / 50 = 20 %.

Expressed in terms of optical rotation (use absolute values):

$$\text{ee} = \frac{\text{specific rotation of mixture}}{\text{specific rotation of pure enantiomer}} \times 100$$

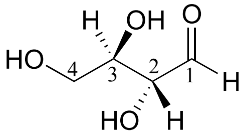
Exercise 3.9: The specific rotation of (*S*)-carvone is (+)61°, measured 'neat' (pure liquid sample, no solvent). The optical rotation of a neat sample of a mixture of *R* and *S* carvone is measured at (-)23°. Which enantiomer is in excess, and what is its ee? What are the percentages of (*R*)- and (*S*)-carvone in the sample?

Chiral molecules are often labeled according to the sign of their specific rotation, as in (S)-(+)-carvone and(R)-(-)-carvone, or (±)-carvone for the racemic mixture. However, there is no relationship whatsoever between a molecule's R/S designation and the sign of its specific rotation. Without performing a polarimetry experiment or looking in the literature, we would have no idea that (-)carvone has the R configuration and (+)-carvone has the S configuration.

Section 3.7: Diastereomers

3.7A: Compounds with multiple stereocenters

We turn our attention next to molecules which have more than one stereocenter. We will start with a common four-carbon sugar called D-erythrose.



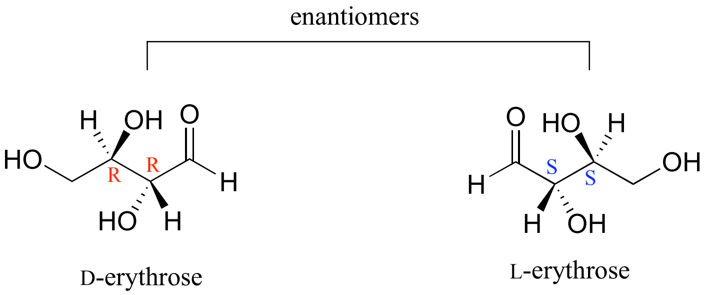
C_2: R
C_3: R

D-erythrose

A note on sugar nomenclature: biochemists use a special system to refer to the stereochemistry of sugar molecules, employing names of historical origin in addition to the designators 'D' and 'L'. You will learn about this system if you take a biochemistry class. We will use the D / L designations here to refer to different sugars, but we won't worry about learning the system.

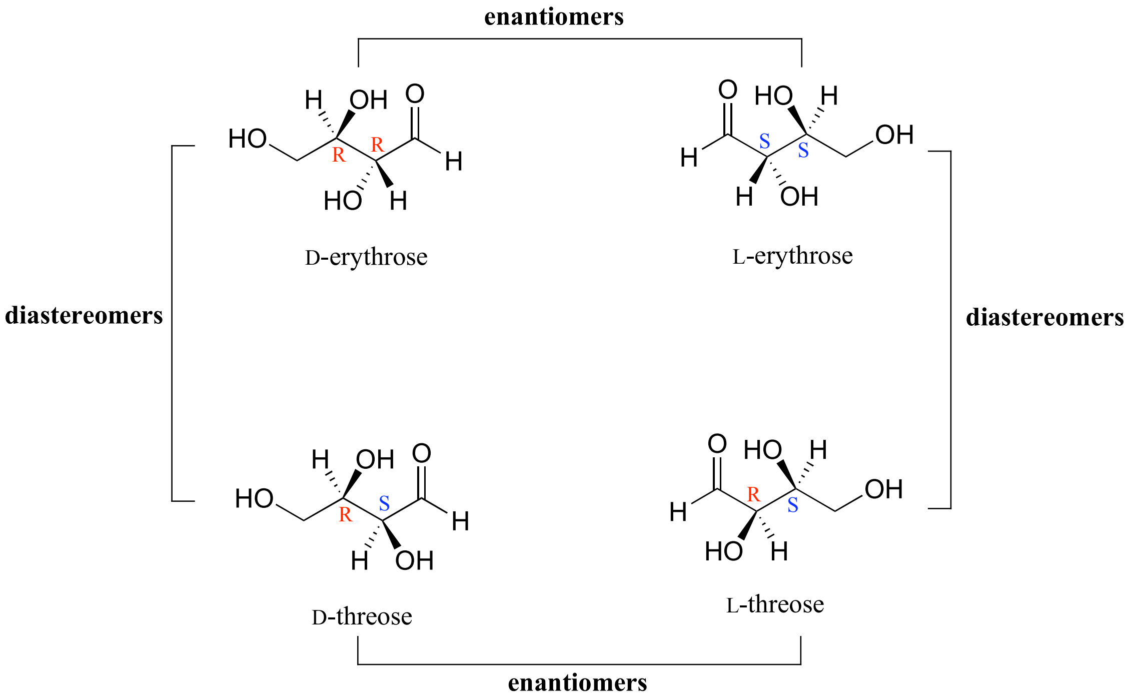
As you can see, D-erythrose is a chiral molecule: C_2 and C_3 are stereocenters, both of which have the R configuration. In addition, you should make a model to convince yourself that it is impossible to find a plane of symmetry through the molecule, regardless of the conformation. Does D-erythrose have an enantiomer? Of course it does – if it is a chiral molecule, it must. The enantiomer of erythrose is its mirror image, and is named

L-erythrose (once again, you should use models to convince yourself that these mirror images of erythrose are not superimposable).



Notice that both stereocenters in L-erythrose both have the S configuration. *In a pair of enantiomers, **all** of the stereocenters are of the opposite configuration.*

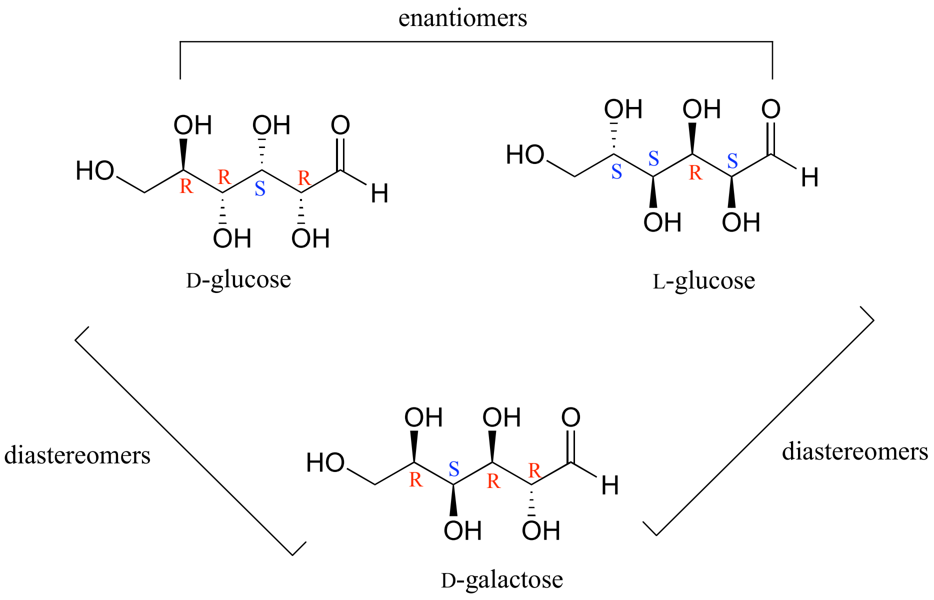
What happens if we draw a stereoisomer of erythrose in which the configuration is S at C_2 and R at C_3? This stereoisomer, which is a sugar called D-threose, is *not* a mirror image of erythrose. D-threose is a **diastereomer** of both D-erythrose and L-erythrose.



The definition of diastereomers is simple: if two molecules are stereoisomers (same molecular formula, same connectivity, different arrangement of atoms in space) but are *not* enantiomers, then they are diastereomers by default. In practical terms, these means that at least one - but not all - of the stereocenters are opposite in the two stereoisomers. By definition, two molecules that are diastereomers are *not* mirror images of each other.

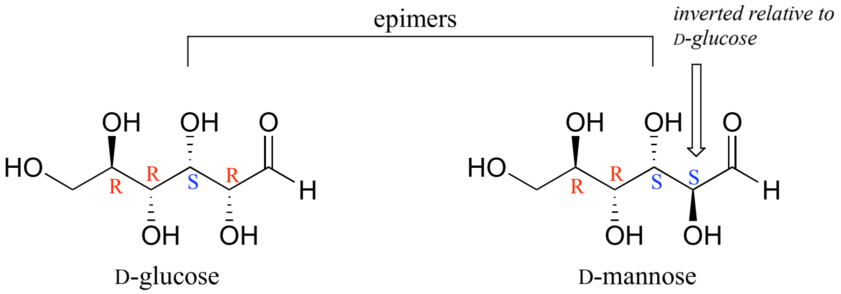
L-threose, the enantiomer of D-threose, has the R configuration at C_2 and the S configuration at C_3. L-threose is a diastereomer of both erythrose enantiomers.

In general, a structure with *n* stereocenters will have 2^n different stereoisomers. (We are not considering, for the time being, the stereochemistry of double bonds – that will come later). For example, let's consider the glucose molecule in its open-chain form (recall that many sugar molecules can exist in either an open-chain or a cyclic form). There are two enantiomers of glucose, called D-glucose and L-glucose. The D-enantiomer is the common sugar that our bodies use for energy. It has $n = 4$ stereocenters, so therefore there are $2^n = 2^4 = 16$ possible stereoisomers (including D-glucose itself).



In L-glucose, all of the stereocenters are inverted relative to D-glucose. That leaves 14 diastereomers of D-glucose: these are molecules in which at least one, but not all, of the stereocenters are inverted relative to D-glucose. One of these 14 diastereomers, a sugar called D-galactose, is shown above: in D-galactose, two of four stereocenters are inverted relative to D-glucose.

D-Mannose is another diastereomer of D-glucose in which the configuration of only one carbon is inverted.

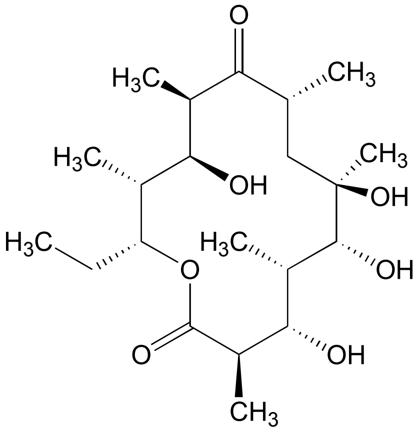


Stereoisomers which differ in only one stereocenter (out of two or more) have a special kind of diastereomeric relationship: they are called **epimers**.

Exercise 3.10: Draw the structure of L-galactose, the enantiomer of D-galactose.

Exercise 3.11: Draw the structure of two more diastereomers of D-glucose. One should be an epimer.

Erythronolide B, a precursor to the 'macrocyclic' antibiotic erythromycin, has 10 stereocenters. It's enantiomer is that molecule in which all 10 stereocenters are inverted.



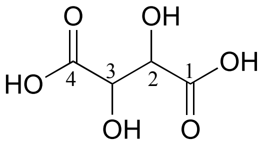
erythronolide B

In total, there are $2^{10} = 1024$ stereoisomers in the erythronolide B family, 1022 of which are diastereomers of the actual structure above.

We know that enantiomers have identical physical properties and equal but opposite degrees of specific rotation. Diastereomers, in theory at least, have different physical properties – we stipulate 'in theory' because sometimes the physical properties of two or more diastereomers are so similar that it is very difficult to separate them. In addition, the specific rotations of diastereomers are unrelated – they could be the same sign or opposite signs, and similar in magnitude or very dissimilar.

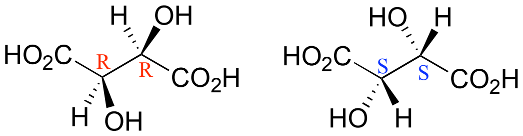
3.7B: *Meso* compounds

It is a general rule that any molecule with at least one stereocenter is chiral – but as with most rules, there is an exception. Some molecules have more than one stereocenter but are actually **a**chiral – these are called ***meso* compounds**. Tartaric acid, a byproduct of the wine-making process, provides a good example.

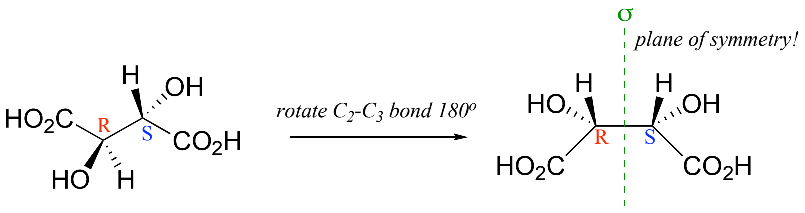


tartaric acid

With two stereocenters, there should, in theory, be four stereoisomers of tartaric acid. In fact, there are only three. First of all, there is a pair of enantiomers with (2R,3R) and (2S, 3S) stereochemistry.



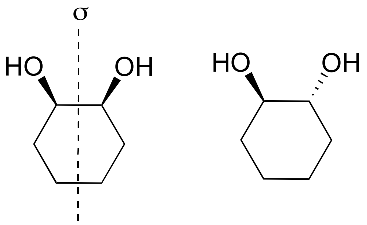
Now, carefully consider a (2S, 3R) stereoisomer. You may notice that, when it is rotated into just the right conformation, this isomer has a plane of symmetry passing through the C_2-C_3 bond.



(meso)-tartaric acid

That means that this molecule is *not* chiral, even though it has two stereocenters! It also means that (2R,3S) tartaric acid and (2S,3R) tartaric acid are not enantiomers, as we might have expected – they are in fact the very same molecule, *meso*-tartaric acid. This achiral molecule is, however, still a diastereomer of both R,R and S,S tartaric acid. Notice that the two 'stereocenters' of (*meso*)-tartaric acid have *the same four substituents* – this is a prerequisite for *meso* compounds; otherwise there would be no plane of symmetry.

Cyclic compounds can also be *meso*. One of many such examples is *cis*-1,2-dihydroxycyclohexane. Note, however, that if the hydroxyl groups are *trans* to each other, the molecule is chiral.

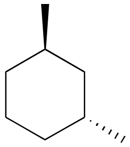


meso chiral

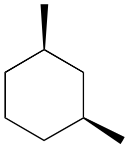
Fortunately for overworked organic chemistry students, the *meso* compound is a very special case, and it is difficult to find many naturally occurring examples .

Exercise 3.12: Which of the molecules shown below are *meso*? Which are chiral? Which are achiral, but not *meso*?

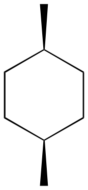
a)



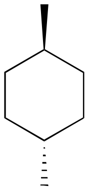
b)



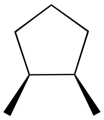
c)



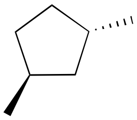
d)



e)



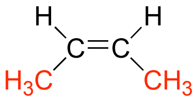
f)



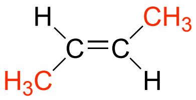
Exercise 3.13: Draw the structure of another dimethylcyclopentane isomer that is *meso* (do not use structures from the previous problem).

3.7C: Stereoisomerism of alkenes

As we saw with *meso* compounds, a molecule or group need not be chiral to have stereoisomers. Asymmetric alkene groups have a defined stereochemistry associated with their structure, but because of their flat geometry they have an inherent plane of symmetry and thus are not chiral. Consider, for example, two stereoisomeric forms of 2-butene.



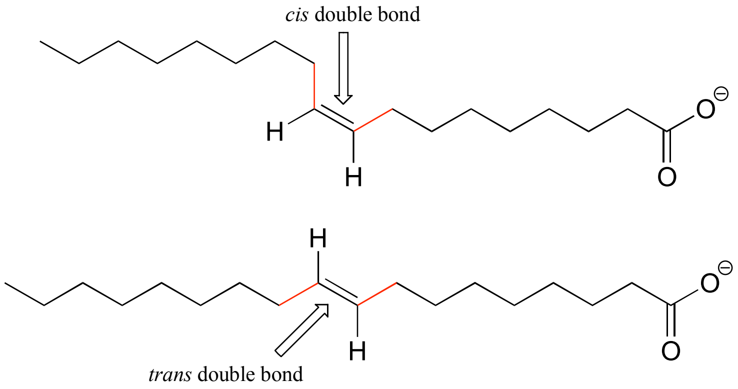
cis-2-butene



trans-2-butene

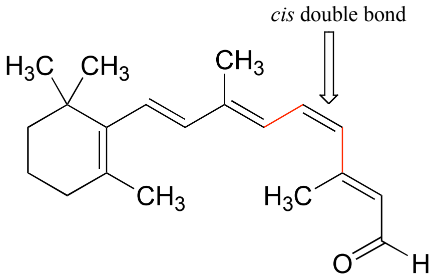
The isomer on the left is *cis*-2-butene, while the isomer on the right is *trans*-2-butene. As we have learned already, the prefix *trans* means 'opposite side', and *cis* means 'same side'. Because there is a barrier to rotation about the C_2-C_3 double bond, the *cis* and *trans* isomers are actually different stereoisomers, not just different conformations of the same molecule. Specifically, they fit the definition of *diastereomers*: they have the same connectivity but a different arrangement of atoms in space, and they are *not* mirror images.

Below are some examples of biomolecules that have diastereomeric alkene groups. Most naturally occurring unsaturated fatty acids have *cis*-double bonds, but *trans*-fatty acids, which are believed to be harmful, are found in margarine and some kinds of meat (see section 16.5D for more about the chemistry of *trans*-fats).

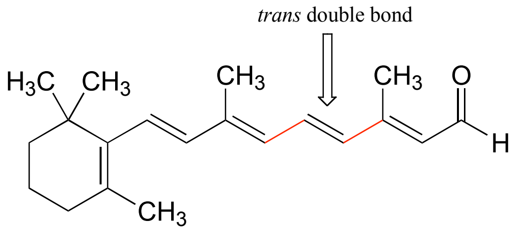


cis and *trans* fatty acids

Retinal is a light-sensitive molecule, derived from vitamin A, that is found in the rod cells of the eye. When light enters the eye through the retina, the 11-*cis* diastereomer of retinal is converted to the all-*trans* diastereomer, changing the shape of the molecule and the way that it binds to the vision protein rhodopsin. This initiates a chain of events that leads to a signal being sent to the vision center of the brain.

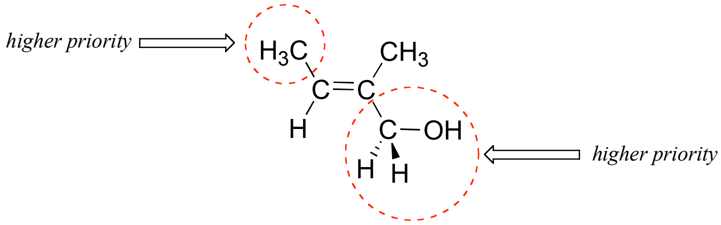


11-*cis*-retinal



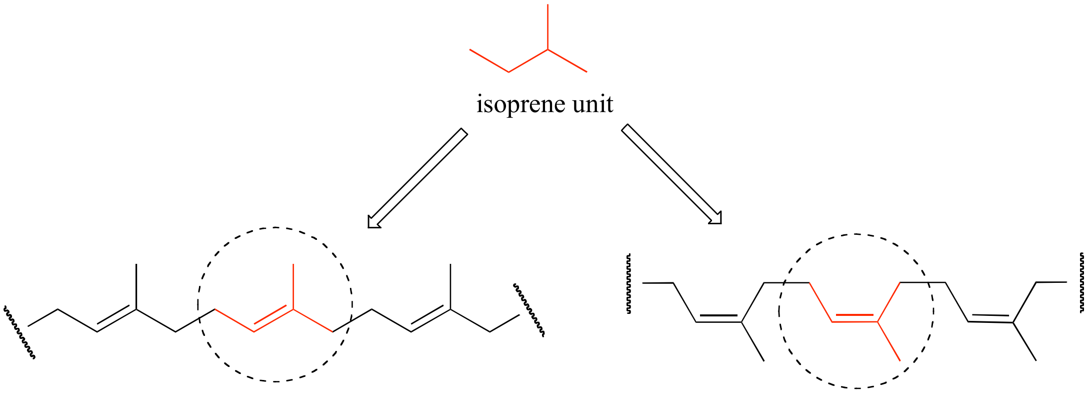
all-*trans*-retinal

While the terms *cis* and *trans* are quite clear in the examples above, in some cases they can be ambiguous, and a more rigorous stereochemical designation is required. To unambiguously designate overall alkene stereochemistry, it is best to use the designators '*E*' and 'Z'. To use this naming system, we first decide which is the higher priority group on each carbon of the double bond, using the same priority rules that we learned for the *R/S* system. If the higher-priority groups are one the same side of the double bond, it is a *Z*-alkene, and if they are on the opposite side it is an *E*-alkene. Shown below is an example of an *E*-alkene: notice that, although the two methyl groups are on the same side (*cis*) relative to one another, the alkene has overall *E* stereochemistry according to the rules of the *E/Z* system because one of the methyl groups takes a lower priority.



an *E*-alkene

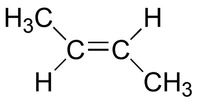
Natural rubber is a polymer composed of five-carbon building blocks, called 'isoprene', that are linked with *Z* stereochemistry. The same isoprene building blocks can also be connected with *E* stereochemistry, leading to a polymer that is a precursor to cholesterol (and many other natural compounds found in all forms of life).



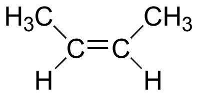
E-isoprenoids
(eg. a precursor to cholesterol)

Z-isoprenoids
(eg. rubber)

As a general rule, alkenes with the bulkiest groups on opposite sides of the double bond are more stable, due to reduced steric strain. The *E* (or *trans*) diastereomer of 2-butene, for example, is more stable than the *Z* (or *cis*) diastereomer.



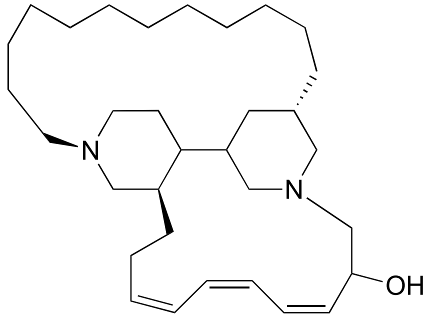
(*E*)-2-butene



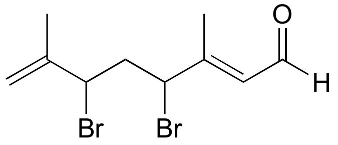
(*Z*)-2-butene

In a reaction where either 2-butene isomer could form, it is the *E* isomer that will predominate (we'll see an example of this phenomenon when we study a reaction called an 'elimination' in section 14.3, and it will also be important when we look at the problem if *trans* fats in section 16.5D).

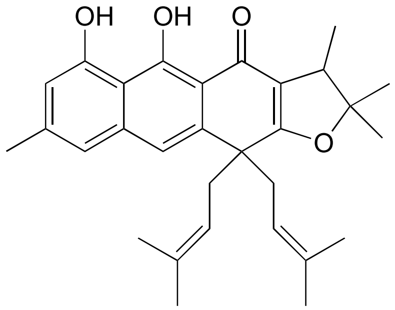
Exercise 3.14: The compounds shown below were all isolated from natural sources and their structures reported in a recent issue of the *Journal of Natural Products*, an American Chemical Society publication. Label all alkene groups that are not inside 5- or 6-membered rings as *E*, *Z*, or *N* (neither E nor Z).



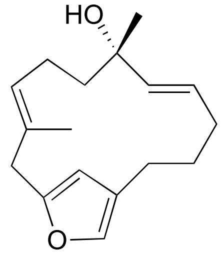
from marine sponge
J. Nat. Prod. **2007**, *70*, 538



from algea
J. Nat. Prod. **2007**, *70*, 596



from Nketto plant in Cameroon
J. Nat. Prod. **2007**, *70*, 600

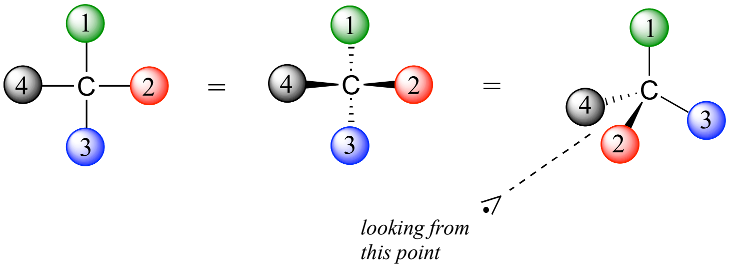


from tree bark in Thailand
J. Nat. Prod. **2007**, *70*, 659

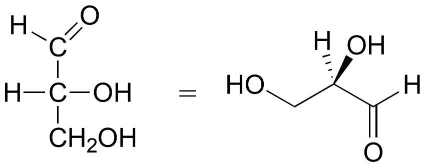
Section 3.8: Fischer and Haworth projections

When reading the chemical and biochemical literature, you are likely to encounter several different conventions for drawing molecules in three dimensions, depending on the context of the discussion. While organic chemists prefer to use the dashed/solid wedge convention to show stereochemistry, biochemists often use drawings called **Fischer projections** and **Haworth projections** to discuss and compare the structure of sugar molecules.

Fisher projections show sugars in their open chain form. In a Fischer projection, the carbon atoms of a sugar molecule are connected vertically by solid lines, while carbon-oxygen and carbon-hydrogen bonds are shown horizontally. Stereochemical information is conveyed by a simple rule: vertical bonds point into the plane of the page, while horizontal bonds point out of the page.

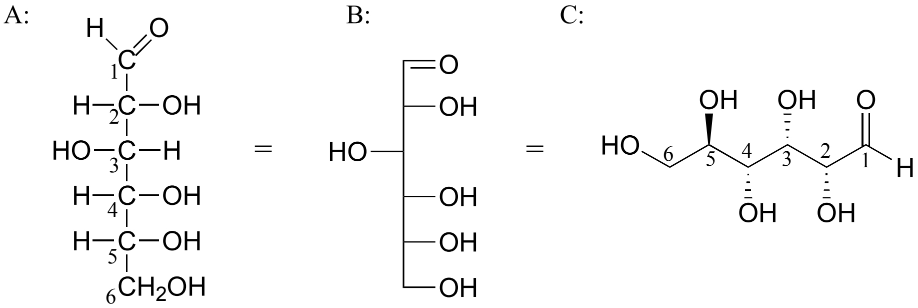


Below are two different representations of (R)-glyceraldehyde, the smallest sugar molecule (also called D-glyceraldehyde in the stereochemical nomenclature used for sugars)

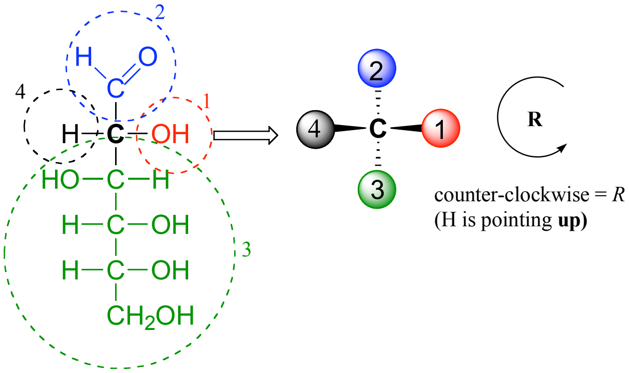


(R)-glyceraldehyde
(D-glyceraldehyde)

Below are three representations of the open chain form of D-glucose: in the conventional Fischer projection (A), in the "line structure" variation of the Fischer projection in which carbons and hydrogens are not shown (B), and finally in the 'zigzag' style (C) that is preferred by organic chemists.

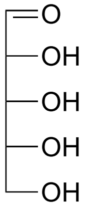


Care must be taken when 'translating' Fischer projection structures into' zigzag' format – it is easy to get the stereochemistry wrong. Probably the best way to make a translation is to simply assign *R*/*S* configurations to each stereocenter, and proceed from there. When deciding whether a stereocenter in a Fischer projection is *R* or *S*, realize that the hydrogen, in a horizontal bond, is pointing *towards* you – therefore, a counterclockwise circle means *R*, and a clockwise circle means *S* (the opposite of when the hydrogen is pointing away from you).

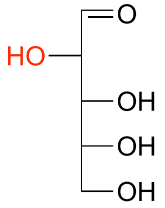


counter-clockwise = *R*
(H is pointing **up)**

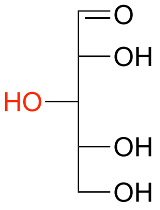
Fischer projections are useful when looking at many different diastereomeric sugar structures, because the eye can quickly pick out stereochemical differences according to whether a hydroxyl group is on the left or right side of the structure.



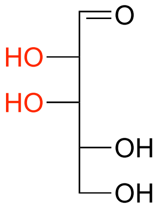
D-ribose



D-arabinose



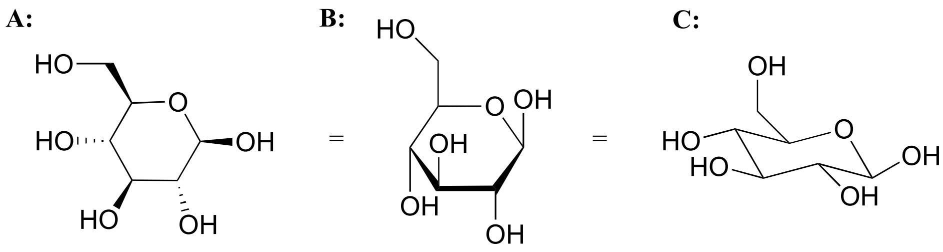
D-xylose



D-lyxose

Exercise 3.15: Draw 'zigzag' structures (using the solid/dash wedge convention to show stereochemistry) for the four sugars in the figure above. Label all stereocenters *R* or *S*.

While Fischer projections are used for sugars in their open-chain form, Haworth projections are often used to depict sugars in their cyclic forms. The β diastereomer of the cyclic form of glucose is shown below in three different depictions, with the Haworth projection in the middle.



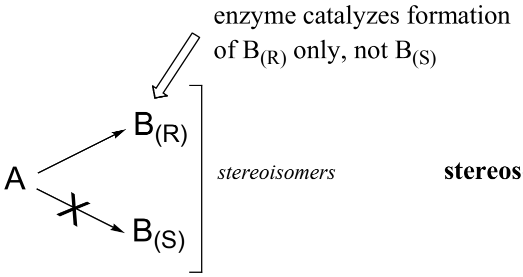
Haworth projection

chair conformation

Notice that although a Haworth projection is a convenient way to show stereochemistry, it does not provide a realistic depiction of conformation. To show *both* conformation and stereochemistry, you must draw the ring in the chair form, as in structure C above.

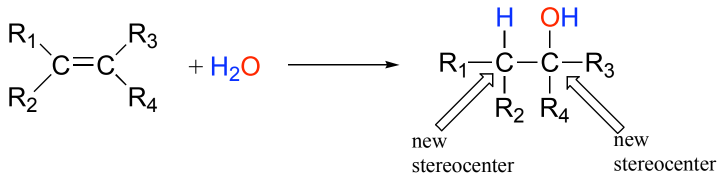
Section 3.9: Stereochemistry and organic reactivity

One of the major differences between laboratory organic reactions (which generally take place free in solution) and biological organic reactions (which generally take place within the very specific, ordered environment of an enzyme) involves the concepts of **stereoselectivity** and **stereospecificity**. In a stereoselective reaction, one stereoisomer is formed preferentially over other possible stereoisomers:

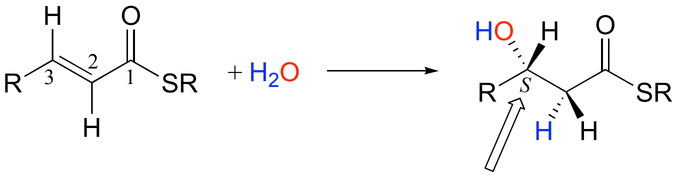

enzyme catalyzes formation of $B_{(R)}$ only, not $B_{(S)}$

stereoselectivity of enzymatic reactions

In chapter 14, we will learn about a reaction type in which a water molecule is 'added' to a double bond.

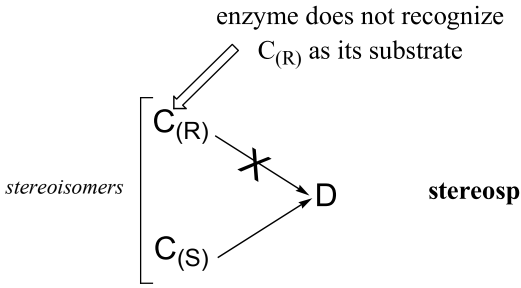


In many cases, this reaction results in the formation of one, or possibly two new stereocenters, depending on the symmetry of the starting alkene double bond. Nonenzymatic laboratory reactions of this type generally are not stereoselective – that is, they result in the formation of mixtures of different stereoisomers. In contrast, a crucial aspect of enzyme-catalyzed biological chemistry is that reactions are almost always highly stereoselective: they result in the formation of only one stereoisomer of their product. For example, this water addition reaction occurs as part of the oxidation of fatty acids:



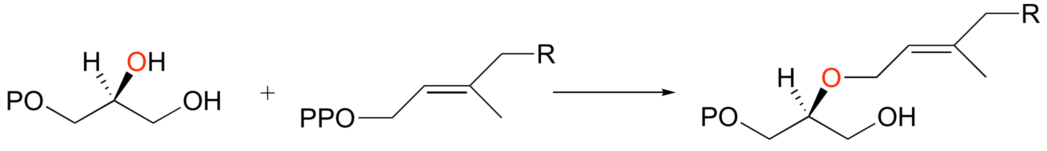
only the *S* enantiomer is formed

Enzymatic reactions are also highly **stereospecific**: this means that an enzyme ‘recognizes’ the stereochemistry of a substrate molecule and only catalyzes its reaction if the substrate stereochemistry is correct.

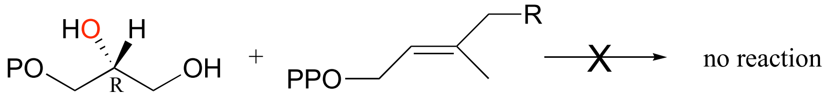


stereospecificity of enzymatic reactions

The enzyme that catalyzes the alkylation of (*S*)-glycerol phosphate, for example, will not work at all with (*R*)-glycerol phosphate.



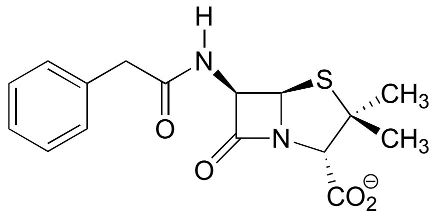
(*S*)-glycerol phosphate



(*R*)-glycerol phosphate

When we begin our study of how organic reactions are catalyzed by enzymes, the reasons for their remarkable stereoselectivity and stereospecificity will become apparent.

After reading the story about thalidomide at the beginning of our discussion of stereochemistry, you have a good appreciation for the importance of stereoisomerism in drug development. It is much safer and more effective if a drug can be provided in stereochemically pure form, without the presence of other ineffective (and possibly dangerous) enantiomers or diastereomers. Drug that are obtained from nature are generally in stereochemically pure form to begin with, because they are synthesized in a living organism by a series of enzymatic reactions. Penicillin, with its three stereocenters and 8 possible stereoisomers, is a good example: as a product synthesized by specific enzymes in *penicillum* mold, it exists as a single stereoisomer.



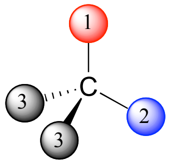
penicillin

Other chiral drugs must be synthesized by humans in the laboratory, where it is much more difficult to produce a single stereoisomer. Medicinal chemists are working very hard to develop new laboratory synthetic techniques which will allow them to better control the stereochemical outcome of their reactions. We will see one example of such a technique in section 16.10B. In addition, many researchers are trying to figure out the enzymatic process by which living things produce useful chiral molecules, so that the enzymes involved may be put to use as *in vitro* synthetic tools. This is important because, while it may be easy to obtain large quantities of *penicillum* mold, many other useful chiral compounds come from species that are difficult, or even impossible, to grow in the lab.

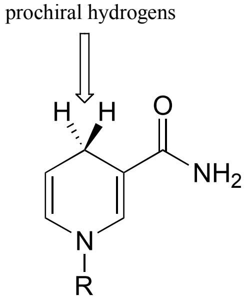
Section 3.10: Prochirality

3.10A: Prochiral substituents on tetrahedral carbons

When a tetrahedral carbon can be converted to a chiral center by changing only one of its attached groups, it is referred to as a '**prochiral'** center. The actual example shown below is the reduced form of a molecule called nicotinamide adenine dinucleotide (NADH), an important participant in many biochemical oxidation/reduction reactions (section 16.4A).

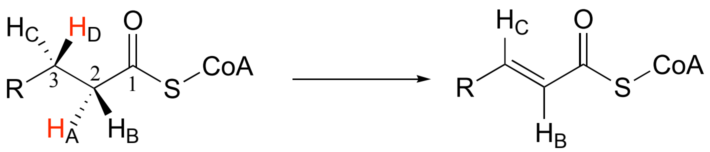


a prochiral carbon center

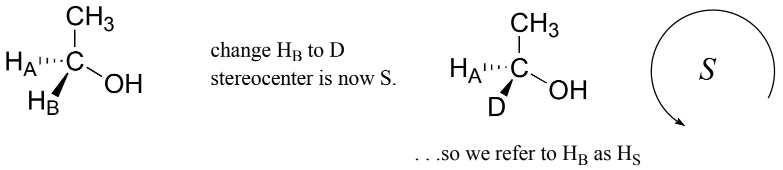


NADH

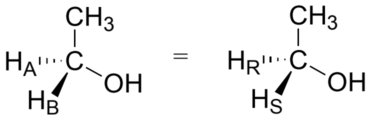
Note that if, in a thought experiment, we changed either one of the indicated hydrogens on NADH to a deuterium (the ^{2}H isotope of hydrogen), the carbon would become a chiral center. Prochirality is an important concept in biological chemistry, because enzymes can distinguish between the two 'identical' groups bound to a prochiral carbon center due to the fact that *they occupy different regions in three-dimensional space*. For example in the following reaction, which is a key step in the oxidation of fatty acids, it is specifically H_A and H_D that are lost, while H_B and H_C remain in the resulting conjugated alkene.



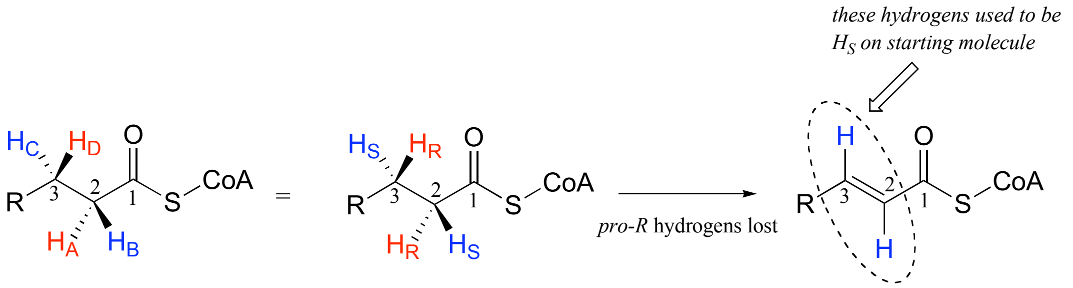
The prochiral hydrogens on C_2 and C_3 of the fatty acid can be designated according to a variation on the *R/S* system. For the sake of clarity, we'll look at a much simpler molecule (ethanol) to explain this system.



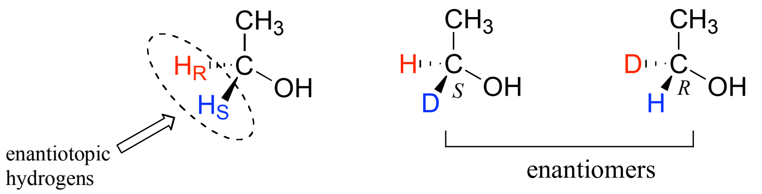
To name the two prochiral hydrogens on ethanol, we again need to engage in a thought experiment. If we, in our imagination, were to arbitrarily change H_B to a deuterium, the molecule would now be chiral and the stereocenter would have the *S* configuration (D has a higher priority than H). For this reason, we can refer to H_B as the *pro-S* hydrogen of ethanol, and label it H_S. Conversely, if we change H_A to D and leave H_B as a hydrogen, the configuration of the molecule becomes *R*, so we can refer to H_A as the *pro-R* hydrogen of ethanol, and label it H_R.



Looking back at our fatty acid example, we see that it is specifically the *pro-R* hydrogens on carbons 2 and 3 that are lost in the reaction.

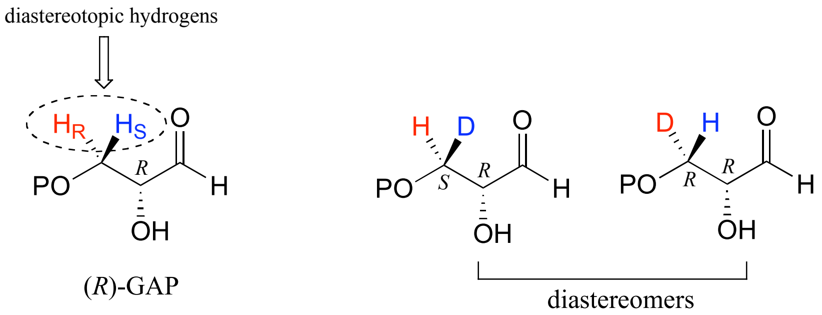


Prochiral hydrogens can be designated either enantiotopic or diastereotopic. If either H_R or H_S on ethanol were replaced by a deuterium, the two resulting molecules would be enantiomers (because there are no other stereocenters)



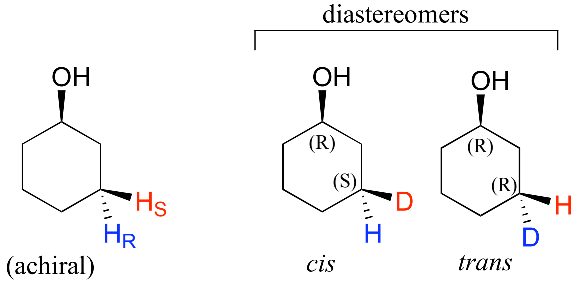
Thus, these two hydrogens are referred to as **enantiotopic**.

In glyceraldehyde-3-phosphate (GAP), however, we see something different:

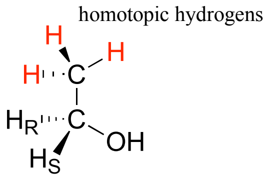


If either H_R or H_S is replaced by a deuterium, the two resulting molecules will be diastereomers - thus, in this molecule, H_R and H_S are referred to as **diastereotopic** hydrogens. The importance of the distinction between enantiotopic and diastereotopic groups will become apparent when we learn about the analytical technique called nuclear magnetic resonance.

Two hydrogens on the same carbon of a substituted ring structure can be diastereotopic - we determine this by carrying out the same thought experiment as discussed above. In the example below, the diastereotopic hydrogens indicated are either on the same side or on the opposite side of the ring relative to the hydroxyl group.



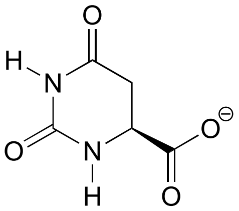
Finally, hydrogens that are completely identical and can be designated neither enantiotopic nor diastereotopic are called **homotopic**. If a homotopic hydrogen is replaced by deuterium, a chiral center is *not* created. The three hydrogen atoms on the methyl (CH_3) group of ethanol (and on *any* methyl group) are homotopic.



Even to an enzyme, all three of these hydrogens will look the same.

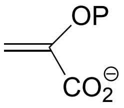
Exercise 3.16: Identify in the molecules below all pairs/groups of hydrogens that are homotopic, enantiotopic, or diastereotopic. When appropriate, label prochiral hydrogens as H_R or H_S.

a)



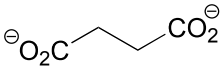
dihydroorotate
(a nucleotide biosynthesis intermediate)

b)



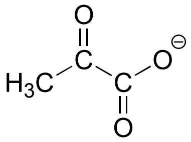
phosphoenolpyruvate
(a glycolysis intermediate)

c)



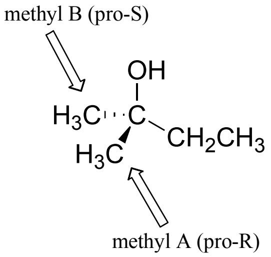
succinate
(a citric acid cycle intermediate)

d)

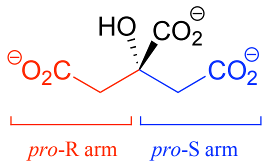


pyruvate
(endpoint of glycolysis)

Groups other than hydrogens can be considered prochiral. The alcohol below has two prochiral methyl groups - methyl A is the *pro-R* methyl, and methyl B is *pro-S*. (How do we make these designations? Simple - just arbitrarily make methyl A higher priority than methyl B, and the compound now has the *R* configuration).

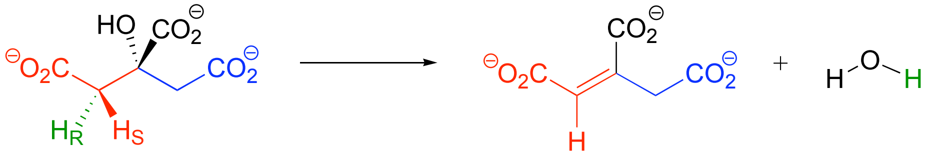


Citrate provides a more complex example. The central carbon is a prochiral center with two 'arms' that are identical except that one can be designated *pro-R* and the other *pro-S*.



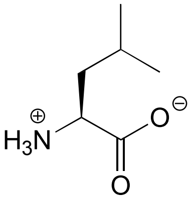
citrate

In a reaction of the citric acid cycle (Krebs cycle), a water molecule is specifically lost on the *pro-R* arm (we will study this reaction in section 14.1B).



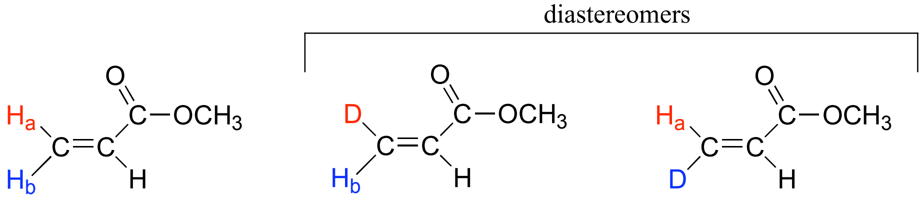
Notice also that it is specifically the *pro-R* hydrogen on the *pro-R* arm of citrate that is lost - one more layer of stereoselectivity!

Exercise 3.17: Assign *pro-R* and *pro-S* designations to all prochiral groups in the amino acid leucine. (*Hint*: there are two pairs of prochiral groups!). Are these prochiral groups diastereotopic or enantiotopic?



leucine

Although an alkene carbon bonded to two identical groups is *not* considered a prochiral center, these two groups *can* be diastereotopic. H_a and H_b on the alkene below, for example, are diastereotopic: if we change one, and then the other, of these hydrogens to deuterium, the resulting compounds are *E* and *Z* diastereomers.

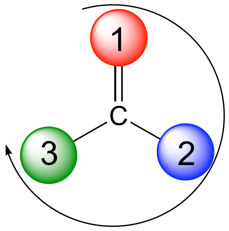


Z-alkene

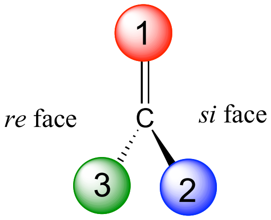
E-alkene

3.10B: Carbonyl and imine carbons as prochiral centers

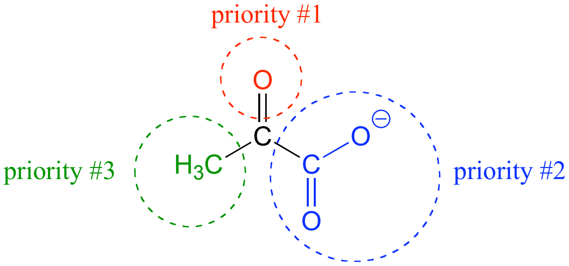
Trigonal planar, sp^2-hybridized carbons are not, as we well know, chiral centers– but they are referred to as prochiral centers if they are bonded to three different substitutuents. As you might expect, we (and the enzymes that catalyze their reactions) can distinguish between the two planar 'faces' of a prochiral sp^2 - hybridized group. These faces are designated by the terms ***re*** and ***si***. To determine which is the *re* and which is the *si* face of a planar organic group, we simply use the same priority rankings that we are familiar with from the R/S system, and trace a circle: *re* is clockwise and *si* is counterclockwise.



looking down on *re* face

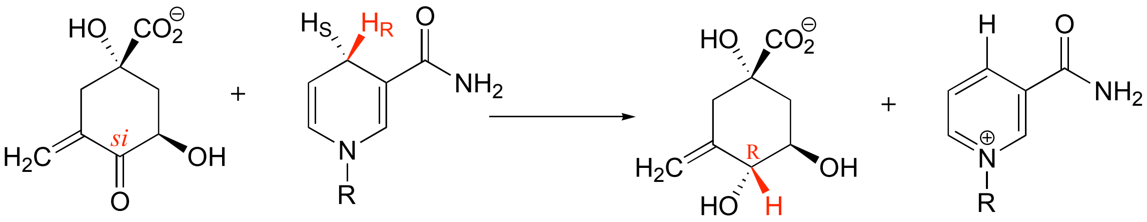


Below, for example, we are looking down on the *re* face of the ketone group in pyruvate:



If we flipped the molecule over, we would be looking at the *si* face of the ketone group. Note that the carboxylate group does not have *re* and *si* faces, because two of the three substituents on that carbon are identical (when the two resonance forms of carboxylate are taken into account).

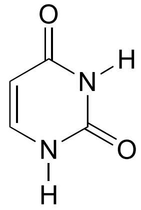
As we will see beginning in chapter 11, enzymes which catalyze reactions at carbonyl carbons act specifically from one side or the other.



pro-R hydrogen is transferred to *si* face of ketone group

Notice that the 'hydrogenation' reaction above is specific not only in terms of which face of the carbonyl group is affected, but also in terms of which of the two diastereotopic hydrogens on NADH is transferred (we will study this type of reaction in more detail in section 16.4).

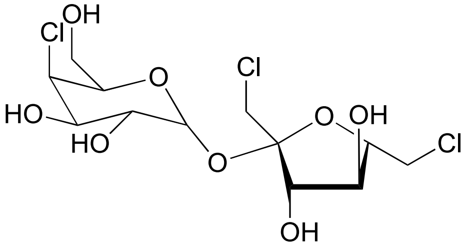
Exercise 3.18: For each of the carbonyl groups in uracil, state whether we are looking at the *re* or the *si* face in the structural drawing below.



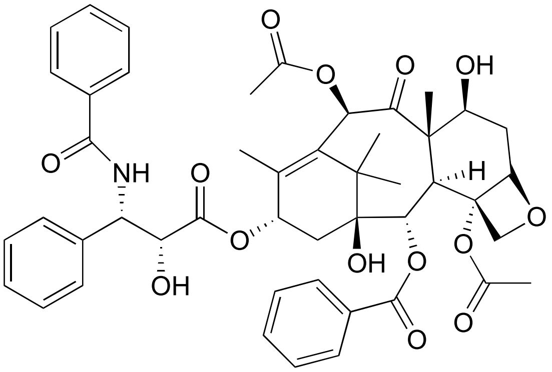
uracil

Chapter 3 problems

P3.1: Below is the structure of sucralose, the artificial sweetener with the brand name Splenda (TM) . Give an *R* or *S* designation to every stereocenter.



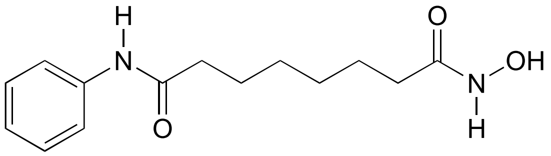
P3.2: Below is the structure of the cancer drug Paclitaxel. Give an *R* or *S* designation to every stereocenter.



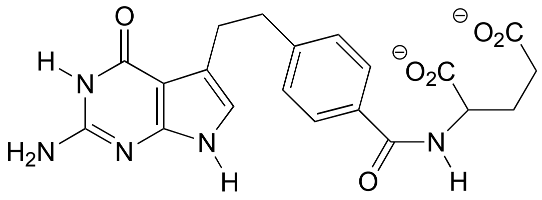
Paclitaxel

P3.3: The four drugs below were featured in a *Chemical & Engineering News* article (April 16, 2007, p. 42) on new drugs that had been developed in university labs.

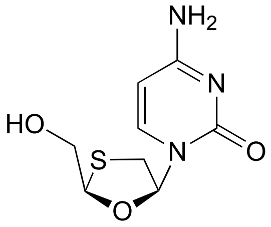
a) Identify each as chiral or achiral, and identify all stereocenters. Also, state how many possible stereoisomers exist for each structure.



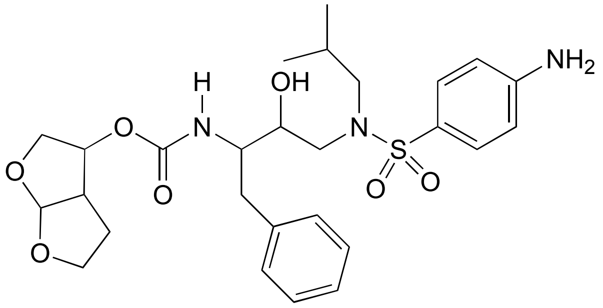
SAHA (suberoylanilide hydroxamic acid)



Alimta

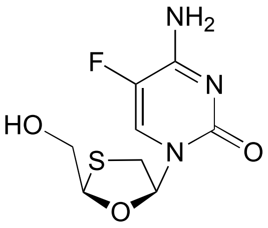


Epivir

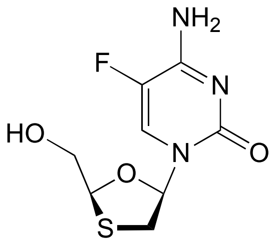


Darunavir

b) Two fluorinated Epivar derivatives (structures A and B below) were also mentioned in this article. What is the relationship between structures A and B?,

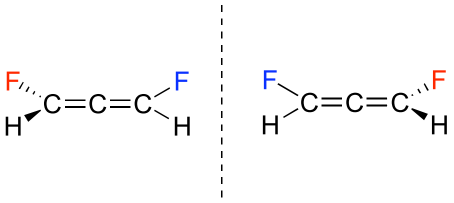


A



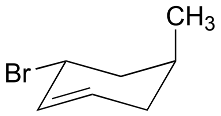
B

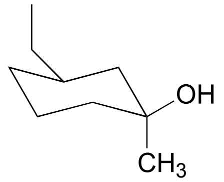
P3.4: What is the relationship between the two structures below?

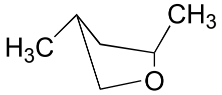


enantiomers of 1,3-difluoroallene

P3.5: Redraw the following structures in the flat ring, solid/dash wedge convention.

a) 

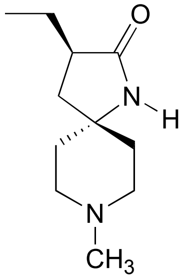
b) 

c) 

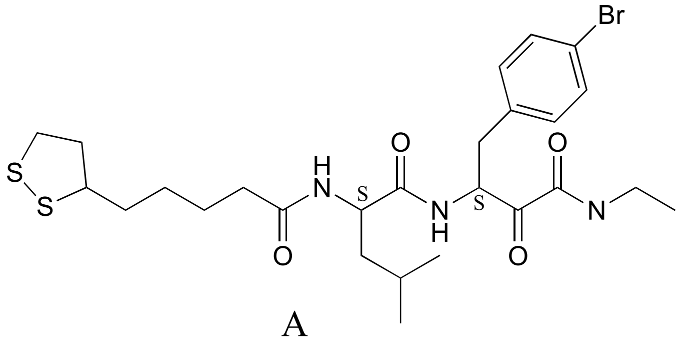
P3.6: Below is an experimental drug for Alzheimer's disease that was mentioned in the March 13, 2007 issue of *Chemical and Engineering News*.

a) Label all stereocenters as *R* or *S*.

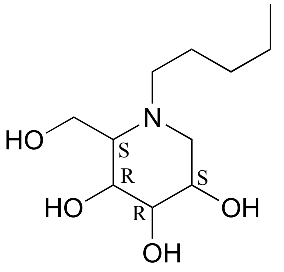
b) Draw the enantiomer of the molecule shown.



P3.7: The molecules below are potential new drugs for the treatment of Duchenne muscular dystrophy (molecule A) and skin cancer (molecule B) (*Chemical and Engineering News* Sept 26, 2005, p. 39). Given the *R*/*S* designations, redraw the structure showing the correct stereochemistry.



A



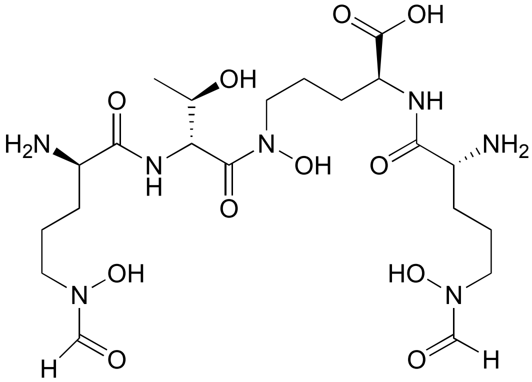
B

P3.8: Draw the structure of the following molecules:

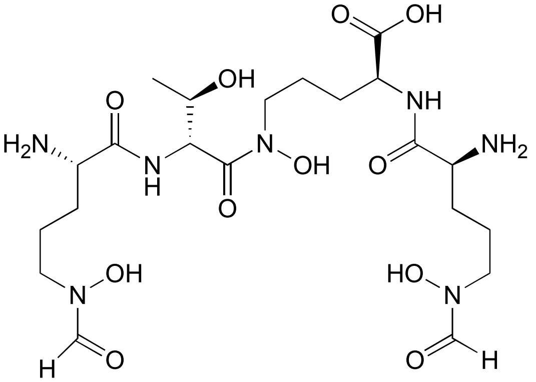
a) (*R*)-3-methyl-3-hexanol
b) (*R*)-1-chloro-1-phenylethane

c) (2*R*, 3*R*)-2,3-dihydroxybutanedioic acid (tartaric acid)
d) (*S*)-(*E*)-4-chloro-3-ethyl-2-pentenoic acid
e) (1*S*, 3*R*)-1-chloro-3-ethylcyclohexane

P3.9: Coelichelin (the structure below to the left) is a natural product from soil bacteria that was identified using a technique known as 'genome mining' (*Chemical and Engineering News* Sept. 19, 2005, p. 11). What is the relationship between coelichelin and the compound shown below and to the right?

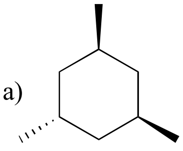


ceolichelin

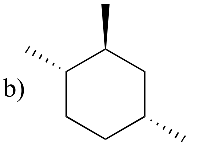


P3.10: Designate the following compounds as chiral or achiral

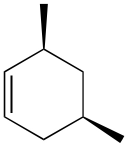
a)



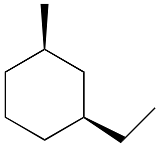
b)



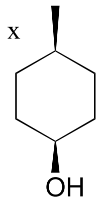
c)



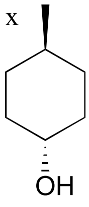
d)



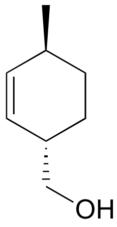
e)



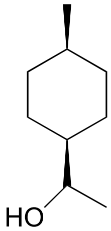
f)



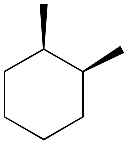
g)



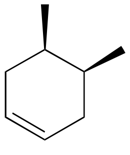
h)



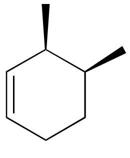
i)



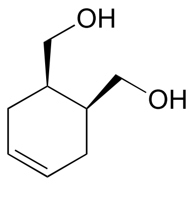
j)



k)

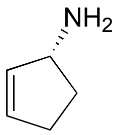


l)

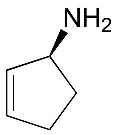


P3.11: Identify the relationships between the following pairs of structures (identical, constitutional isomers, enantiomers, diastereomers/not epimers, epimers, or completely different compounds).

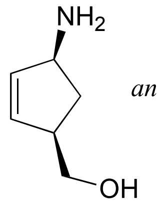
a)



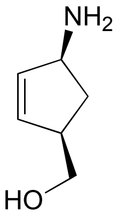
and



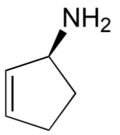
b)



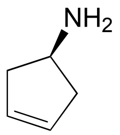
and



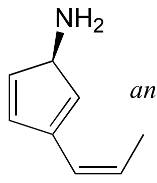
c)



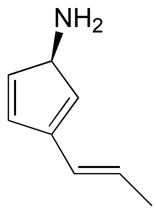
and



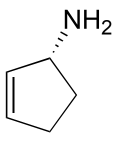
d)



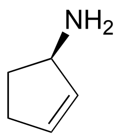
and



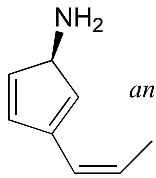
e)



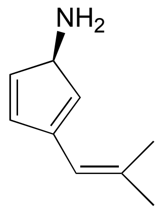
and



f)



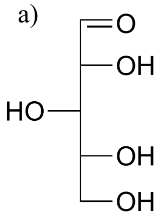
and



P3.12:

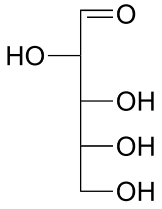
Identify the relationships between each of the following pairs of pentose sugar molecules (identical, constitutional isomers, enantiomers, diastereomers/not epimers, epimers, or completely different compounds). State how many stereocenters each sugar contains.

a)



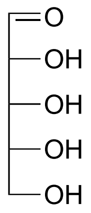
D-xylose

and



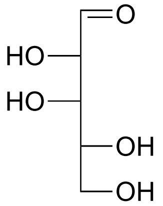
D-arabinose

b)



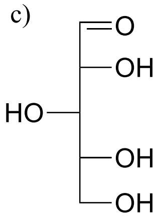
D-ribose

and



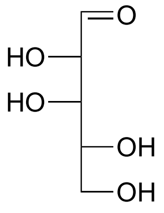
D-lyxose

c)



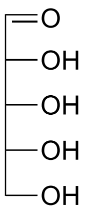
D-xylose

and



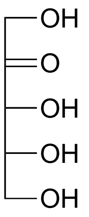
D-lyxose

d)



D-ribose

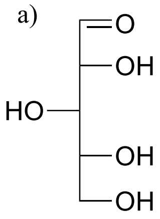
and



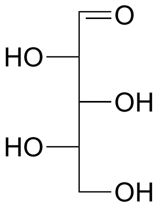
D-ribulose

P3.13: Identify the relationships between each of the following pairs of pentose sugar molecules (identical, constitutional isomers, enantiomers, diastereomers/not epimers, epimers, or completely different compounds).

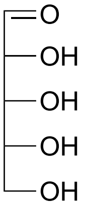
a)



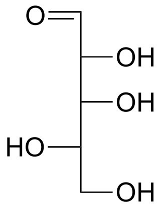
and



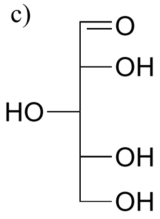
b)



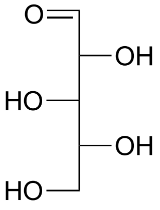
and



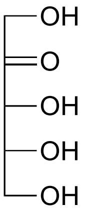
c)



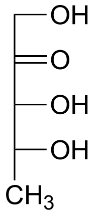
and



d)



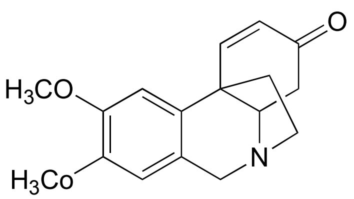
and



P3.14: A novel and very efficient laboratory synthesis of the natural alkaloid oxomaritidine was described in the March 6, 2006 volume *of Chemical and Engineering News* (p. 17). The structure is shown below in two dimensions.

a) Identify any chiral carbons in the molecule.

b) Identify any prochiral hydrogens in the molecule

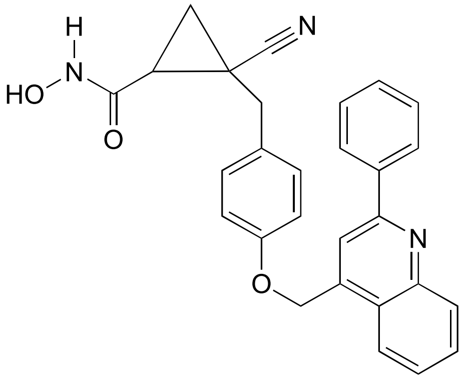


oxomaritidine

P3.15: The compound drawn below (not showing stereochemistry) has been identified as a potential anti-inflammatory agent by scientists at Schering-Plough a pharmaceutical company (see *Chemical and Engineering News* Nov. 28, 2005 p. 29).

a) How many stereoisomers are possible for the compound?

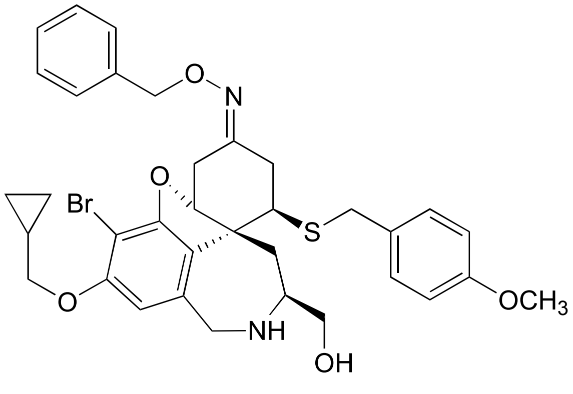
b) Identify all prochiral hydrogens in the compound.



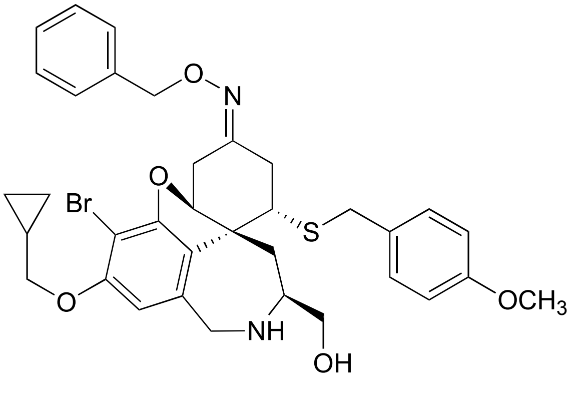
P3.16: Secramine is a synthetic compound that has been shown to interfere with the transport of newly synthesized proteins in the cell (see *Chemical and Engineering News* Nov. 28, 2005, p. 27). Also drawn below is a (hypothetical) isomer of secramine.

a) Identify the relationship between the two isomers: are they consitutional isomers, confomational isomers, enantiomers, or diastereomers?

b) Locate a five-membered ring in the secramine structure.

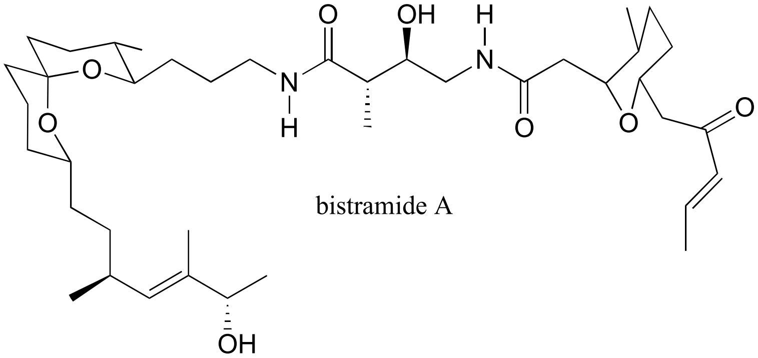


secramine



isomer of secramine

P3.17: The natural product bistramide A has been shown to bind to actin, an important structural protein in the cell, and supress cell proliferation (see *Chemical and Engineering News* Nov. 21, 2005, p. 10).

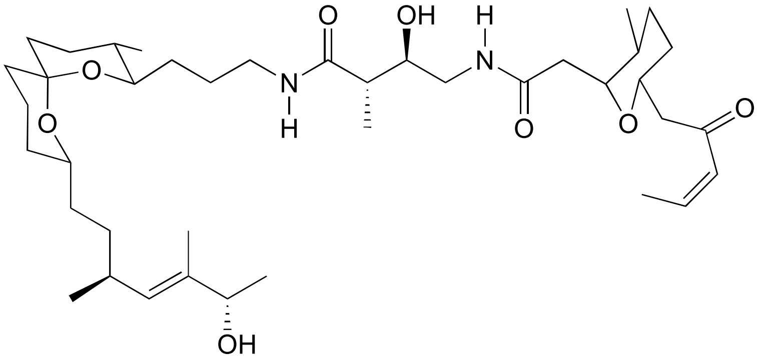


bistramide A

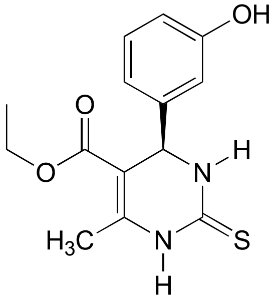
a) Label the alkene functional groups as *E*, *Z*, or *N* (no *E*/*Z* designation possible)

b) Theoretically, how many stereoisomers are possible for bistramide A?

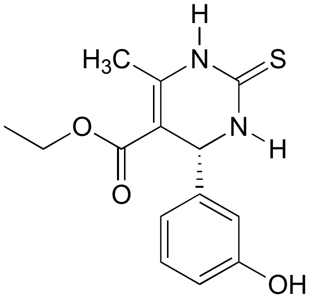
c) What is the relationship between bistramide A and the structure below?



P3.18: Structure A below is that of a newly-developed chiral catalyst (see *Chemical and Engineering News* Nov. 21, 2005 p. 50). What is the relationship between structures A and B?



structure A



structure B

P3.19:

a) Draw Newman projections of the *gauche* and the *anti* conformations of 1,2-ethanediol.

b) Why might the *gauche* conformation be expected to be the more stable of the two?

c) Do you think that *gauche* is also the most stable conformation of 1,2-dimethoxyethane? Explain.

P3.20: Draw a Newman projection of the highest energy conformation of 2-methylbutane, looking down the C_2-C_3 bond.

P3.21: Draw the lower energy chair conformation of *trans*-1,2-dimethylcyclohexane.

P3.22: Draw the lower energy chair conformation of *cis*-1-ethyl-2-methylcyclohexane

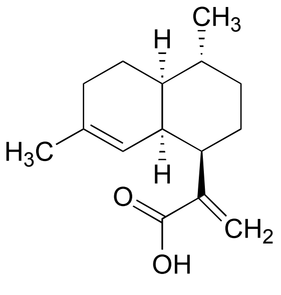
P3.23: Draw the lower-energy chair conformations of *cis*-1,4-diethylcyclohexane and *trans*-1,4-diethylcyclohexane. Which is expected to be lowest in energy?

P3.24: Draw the lower-energy chair conformations of *cis*-1,3-diethylcyclohexane and *trans*-1,3-diethylcyclohexane. Which is expected to be lowest in energy?

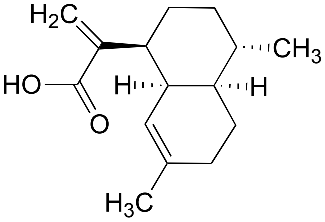
P3.25: Draw the chair conformation of *cis*-1,2-dimethylcyclohexane. In this conformation, does the molecule have a mirror plane? Draw a boat conformation of the same molecule that contains a mirror plane, and draw a line to depict that plane (models will help).

P3.26: Structure A below is that of artemisinic acid, a synthetic precursor to the anti-malaria drug artemisinin. Identify the relationship (identical, completely different, constitutional isomers, enantiomers, or diastereomers) between the following pairs of structures (it will be *very* helpful to use models!)

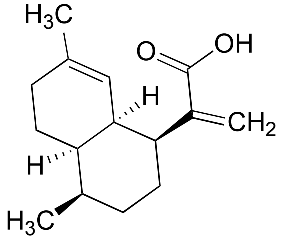
a) A and B b) A and C c) B and C d) B and D



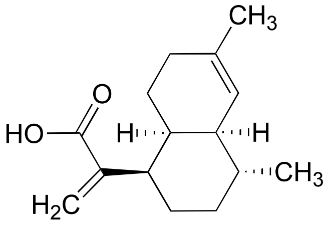
A



B

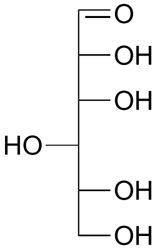


C

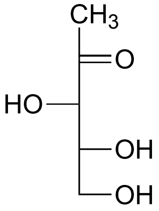


D

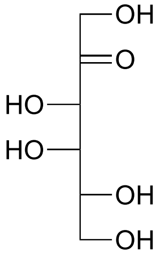
P3.27: Redraw the following sugar structures using the zig-zag, solid/dash wedge convention, and assign *R*/*S* designations to all stereocenters.



D-gulose



D-1-deoxyxylulose



D-tagatose

Chapter 4

Structure determination, part I:

Infrared spectroscopy, UV-visible spectroscopy, and mass spectrometry

Introduction

In the first three chapters of this text, we have focused our efforts on learning about the structure of organic compounds. Now that we know what organic molecules look like, we can start to address the question of *how* chemists are able to elucidate organic structures. The individual atoms and functional groups in organic compounds are far too small to be directly observed or photographed, even with the best electron microscope. How, then, are chemists able to draw with confidence the bonding arrangements in organic molecules, even simple ones such as acetone or ethanol?

The answer lies, for the most part, in a field of chemistry called **molecular spectroscopy**. Spectroscopy is the study of how electromagnetic radiation, across a spectrum of different wavelengths, interacts with molecules - and how these interactions can be quantified, analyzed, and ultimately interpreted to gain information about molecular structure.

After first reviewing some basic information about the properties of light and introducing the basic ideas behind spectroscopy, we will move to a discussion of infrared (IR) spectroscopy, a technique which is used in organic chemistry to detect the presence or absence of common functional groups. Next, we will look at ultraviolet-visible (UV-vis) spectroscopy, in which light of a shorter wavelength is employed to provide information about organic molecules containing conjugated π-bonding systems.

In the final section of this chapter, we will change tack slightly and consider another analytical technique called mass spectrometry (MS). Here, we learn about the structure of a molecule by, in a sense, taking a hammer to it and smashing it into small pieces, then measuring the mass of each piece. Although this metaphorical description makes mass spectrometry sound somewhat crude, it is in fact an extremely powerful and sensitive technique, one which has in recent years become central to the study of life at the molecular level.

Looking ahead, Chapter 5 will be devoted to the study of nuclear magnetic resonance (NMR) spectroscopy, where we use ultra-strong magnets and radio frequency radiation to learn about the electronic environment of individual atoms in a molecule. For most organic chemists, NMR is the single most powerful analytical tool available in terms of the wealth of detailed information it can provide about the structure of a molecule. It is the closest thing we have to a 'molecular camera'.

In summary, the analytical techniques we will be studying in this chapter and the next primarily attempt to address the following questions about an organic molecule:

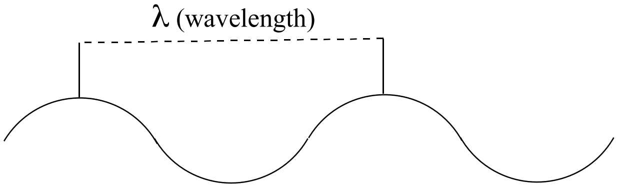
> Infrared (IR) spectroscopy: *what functional groups does the molecule contain?*
> Ultraviolet-visible (UV-vis) spectroscopy: *To what extent does the molecule contain a system of conjugated π bonds?*
> Mass spectrometry (MS): *What is the atomic weight of the molecule and its common fragments?*
> Nuclear magnetic resonance spectroscopy (NMR): *What is the overall bonding framework of the molecule?*

Section 4.1: Introduction to molecular spectroscopy

4.1A: The electromagnetic spectrum

Electromagnetic radiation, as you may recall from a previous chemistry or physics class, is composed of electrical and magnetic waves which oscillate on perpendicular planes. Visible light is electromagnetic radiation. So are the gamma rays that are emitted by spent nuclear fuel, the x-rays that a doctor uses to visualize your bones, the ultraviolet light that causes a painful sunburn when you forget to apply sun block, the infrared light that the army uses in night-vision goggles, the microwaves that you use to heat up your frozen burritos, and the radio-frequency waves that bring music to anybody who is old-fashioned enough to still listen to FM or AM radio.

Just like ocean waves, electromagnetic waves travel in a defined direction. While the speed of ocean waves can vary, however, the speed of electromagnetic waves – commonly referred to as the speed of light – is essentially a constant, approximately 300 million meters per second. This is true whether we are talking about gamma radiation or visible light. Obviously, there is a big difference between these two types of waves – we are surrounded by the latter for more than half of our time on earth, whereas we hopefully never become exposed to the former to any significant degree. The different properties of the various types of electromagnetic radiation are due to differences in their wavelengths, and the corresponding differences in their energies: *shorter wavelengths correspond to higher energy*.



High-energy radiation (such as gamma- and x-rays) is composed of very short waves – as short as 10^{-16} meter from crest to crest. Longer waves are far less energetic, and thus are less dangerous to living things. Visible light waves are in the range of 400 – 700 nm (nanometers, or 10^{-9} m), while radio waves can be several hundred meters in length.

The notion that electromagnetic radiation contains a quantifiable amount of energy can perhaps be better understood if we talk about light as a stream of *particles*, called **photons**, rather than as a wave. (Recall the concept known as 'wave-particle duality': at the quantum level, wave behavior and particle behavior become indistinguishable, and very small particles have an observable 'wavelength'). If we describe light as a stream of photons, the energy of a particular wavelength can be expressed as:

equation 4.1 $$E = hc/\lambda$$

where E is energy in kcal/mol, λ (the Greek letter *lambda*) is wavelength in meters, c is 3.00 x 10^8 m/s (the speed of light), and h is 9.537 x 10^{-14} kcal·s·mol^{-1}, a number known as Planck's constant.

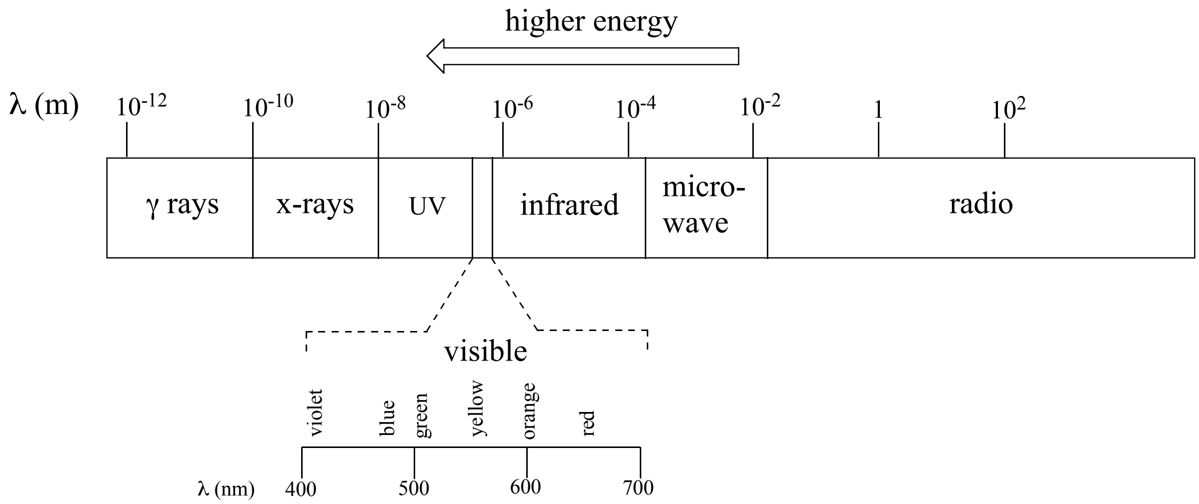
Because electromagnetic radiation travels at a constant speed, each wavelength corresponds to a given frequency, which is the number of times per second that a crest passes a given point. Longer waves have lower frequencies, and shorter waves have higher frequencies. Frequency is commonly reported in hertz (Hz), meaning 'cycles per second', or 'waves per second'. The standard unit for frequency is s^{-1}.

When talking about electromagnetic waves, we can refer either to wavelength or to frequency - the two values are interconverted using the simple expression:

equation 4.2 $$\lambda\upsilon = c$$

where υ (the Greek letter '*nu*') is frequency in s^{-1}. Visible red light with a wavelength of 700 nm, for example, has a frequency of 4.29 x 10^{14} Hz, and an energy of 40.9 kcal per mole of photons.

The full range of electromagnetic radiation wavelengths is referred to as the **electromagnetic spectrum**.



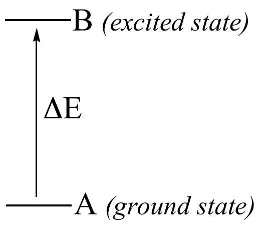
Notice in the figure above that visible light takes up just a narrow band of the full spectrum. White light from the sun or a light bulb is a mixture of all of the visible wavelengths. You see the visible region of the electromagnetic spectrum divided into its different wavelengths every time you see a rainbow: violet light has the shortest wavelength, and red light has the longest.

Exercise 4.1: Visible light has a wavelength range of about 400-700 nm. What is the corresponding frequency range? What is the corresponding energy range, in kcal/mol of photons?

4.1B: Molecular spectroscopy – the basic idea

In a spectroscopy experiment, electromagnetic radiation of a specified range of wavelengths is allowed to pass through a sample containing a compound of interest. The sample molecules absorb energy from some of the wavelengths, and as a result jump from a low energy 'ground state' to some higher energy 'excited state'. Other wavelengths are *not* absorbed by the sample molecule, so they pass on through. A detector on the other side of the sample records which wavelengths were absorbed, and to what extent they were absorbed.

Here is the key to molecular spectroscopy: *a given molecule will specifically absorb only those wavelengths which have energies that correspond to the energy difference of the transition that is occurring*. Thus, if the transition involves the molecule jumping from ground state A to excited state B, with an energy difference of ΔE, the molecule will specifically absorb radiation with wavelength that corresponds to ΔE, while allowing other wavelengths to pass through unabsorbed.

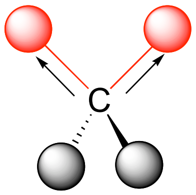


By observing which wavelengths a molecule absorbs, and to what extent it absorbs them, we can gain information about the nature of the energetic transitions that a molecule is able to undergo, and thus information about its structure.

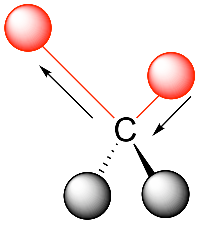
These generalized ideas may all sound quite confusing at this point, but things will become much clearer as we begin to discuss specific examples.

Section 4.2: Infrared spectroscopy

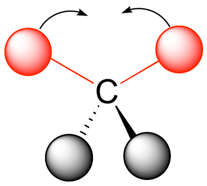
Covalent bonds in organic molecules are not rigid sticks – rather, they behave more like springs. At room temperature, organic molecules are always in motion, as their bonds stretch, bend, and twist. These complex vibrations can be broken down mathematically into individual **vibrational modes**, a few of which are illustrated below.



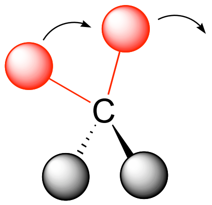
symmetric stretching



asymmetric stretching



scissoring

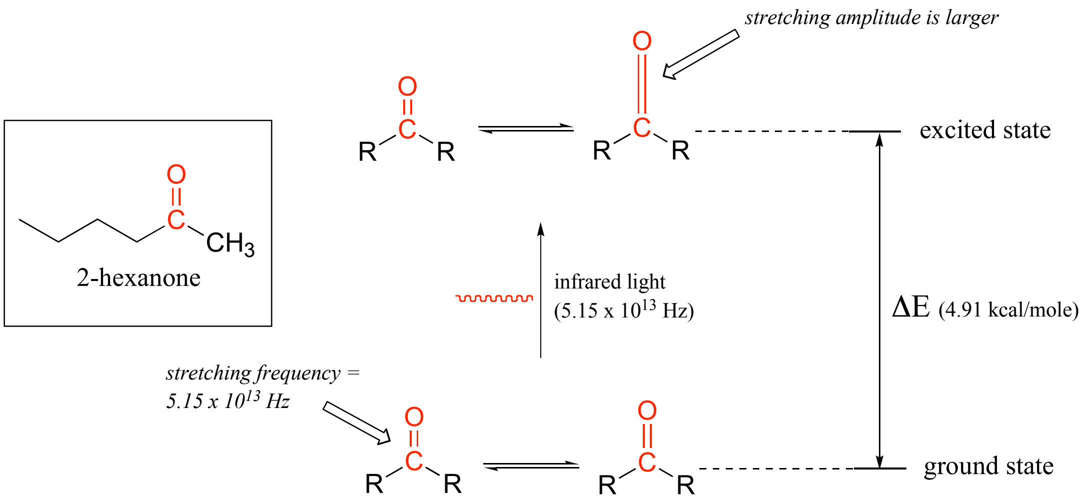


rocking

The energy of molecular vibration is *quantized* rather than continuous, meaning that a molecule can only stretch and bend at certain 'allowed' frequencies. If a molecule is exposed to electromagnetic radiation that matches the frequency of one of its vibrational modes, it will in most cases absorb energy from the radiation and jump to a higher vibrational energy state - what this means is that the *amplitude* of the vibration will increase, but the vibrational *frequency* will remain the same. The difference in energy between the two vibrational states is equal to the energy associated with the wavelength of radiation that was absorbed. It turns out that it is the *infrared* region of the electromagnetic spectrum which contains frequencies corresponding to the vibrational frequencies of organic bonds.

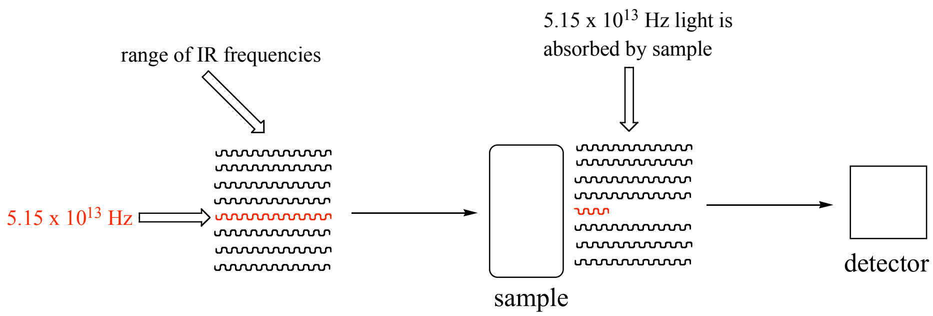
Let's take 2-hexanone as an example. Picture the carbonyl bond of the ketone group as a spring. This spring is constantly bouncing back and forth, stretching and compressing, pushing the carbon and oxygen atoms further apart and then pulling them together. This is the **stretching mode** of the carbonyl bond. In the space of one second, the spring 'bounces' back and forth 5.15 x 10^{13} times - in other words, the ground-state frequency of carbonyl stretching for a the ketone group is about 5.15 x 10^{13} Hz.

If our ketone sample is irradiated with infrared light, the carbonyl bond will specifically absorb light with this same frequency, which by equations 4.1 and 4.2 corresponds to a wavelength of $\lambda = 5.83 \times 10^{-6}$ m and an energy of 4.91 kcal/mol. When the carbonyl bond absorbs this energy, it jumps up to an excited vibrational state.



The value of ΔE - the energy difference between the low energy (ground) and high energy (excited) vibrational states - is equal to 4.91 kcal/mol, the same as the energy associated with the absorbed light frequency. The molecule does not remain in its excited vibrational state for very long, but quickly releases energy to the surrounding environment in form of heat, and returns to the ground state.

With an instrument called an infrared spectrophotometer, we can 'see' this vibrational transition. In the spectrophotometer, infrared light with frequencies ranging from about 10^{13} to 10^{14} Hz is passed though our sample of cyclohexane. Most frequencies pass right through the sample and are recorded by a detector on the other side.



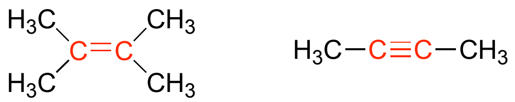
Our 5.15 x 10^{13} Hz carbonyl stretching frequency, however, is absorbed by the 2-hexanone sample, and so the detector records that the intensity of this frequency, after having passed through the sample, is something less than 100% of its initial intensity.

The vibrations of a 2-hexanone molecule are not, of course, limited to the simple stretching of the carbonyl bond. The various carbon-carbon bonds also stretch and bend, as do the carbon-hydrogen bonds, and all of these vibrational modes also absorb different frequencies of infrared light.

The power of infrared spectroscopy arises from the observation that *different functional groups have different characteristic absorption frequencies*. The carbonyl bond in a ketone, as we saw with our 2-hexanone example, typically absorbs in the range of 5.11 - 5.18 x 10^{13} Hz, depending on the molecule. The carbon-carbon triple bond of an alkyne, on the other hand, absorbs in the range 6.30 - 6.80 x 10^{13} Hz. The technique is therefore very useful as a means of identifying which functional groups are present in a molecule of interest. If we pass infrared light through an unknown sample and find that it absorbs in the carbonyl frequency range but not in the alkyne range, we can infer that the molecule contains a carbonyl group but not an alkyne.

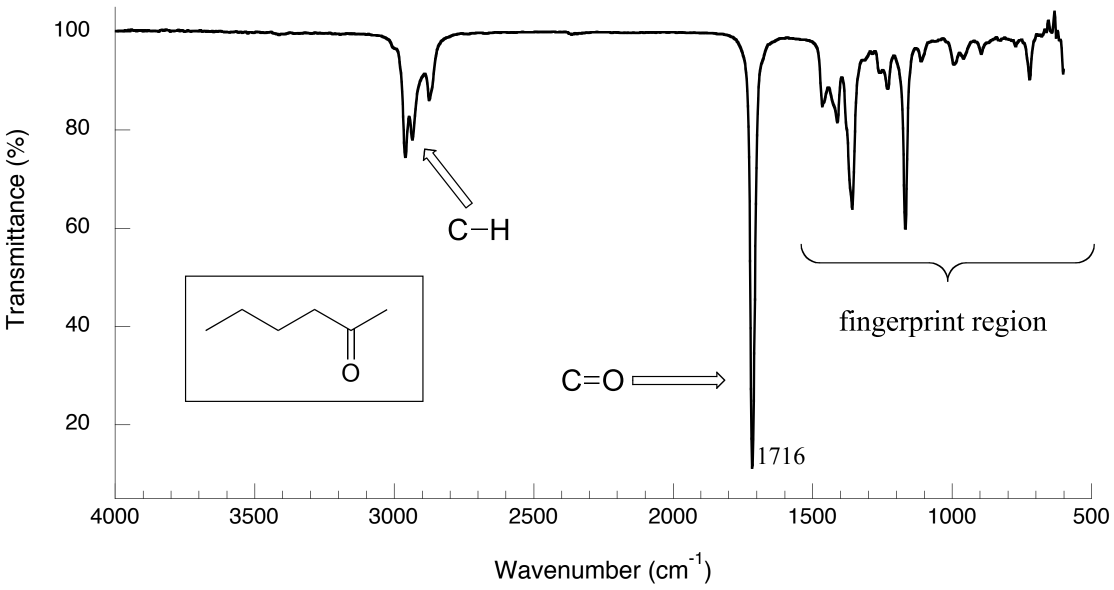
Some bonds absorb infrared light more strongly than others, and some bonds do not absorb at all. *In order for a vibrational mode to absorb infrared light, it must result in a periodic change in the dipole moment of the molecule*. Such vibrations are said to be **infrared active**. In general, the greater the polarity of the bond, the stronger its IR absorption. The carbonyl bond is very polar, and absorbs very strongly. The carbon-carbon triple bond in most alkynes, in contrast, is much less polar, and thus a stretching vibration does not result in a large change in the overall dipole moment of the molecule. Alkyne groups absorb rather weakly compared to carbonyls.

Some kinds of vibrations are **infrared inactive**. The stretching vibrations of completely symmetrical double and triple bonds, for example, do not result in a change in dipole moment, and therefore do not result in any absorption of light (but other bonds and vibrational modes in these molecules *do* absorb IR light).



infrared-inactive double and triple bonds

Now, let's look at some actual output from IR spectroscopy experiments. Below is the IR spectrum for 2-hexanone.



There are a number of things that need to be explained in order for you to understand what it is that we are looking at. On the horizontal axis we see IR wavelengths expressed in terms of a unit called **wavenumber** (cm^{-1}), which tells us how many waves fit into one centimeter. On the vertical axis we see '% transmittance', which tells us how strongly light was absorbed at each frequency (100% transmittance means no absorption occurred at that frequency). The solid line traces the values of % transmittance for every wavelength – the 'peaks' (which are actually pointing down) show regions of strong absorption. For some reason, it is typical in IR spectroscopy to report wavenumber values rather than wavelength (in meters) or frequency (in Hz). The 'upside down' vertical axis, with absorbance peaks pointing down rather than up, is also a curious convention in IR spectroscopy. We wouldn't want to make things too easy for you!

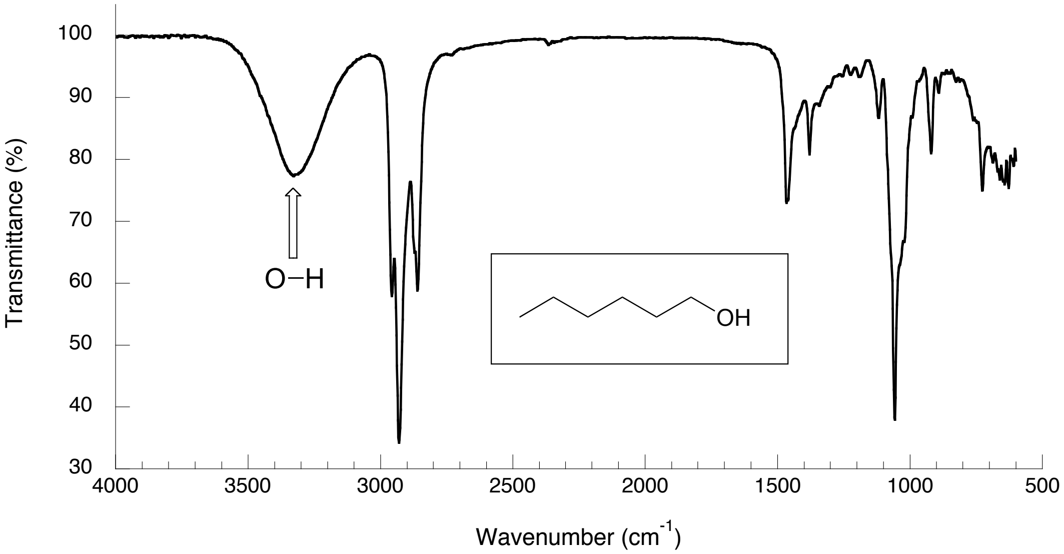
Exercise 4.2: Express the wavenumber value of 3000 cm^{-1} in terms of wavelength (in meter units).

The key absorption peak in this spectrum is that from the carbonyl double bond, at 1716 cm^{-1} (corresponding to a wavelength of 5.86 μm, a frequency of 5.15×10^{13} Hz, and a ΔE value of 4.91 kcal/mol). Notice how strong this peak is, relative to the others on the spectrum: *a strong peak in the 1650-1750 cm^{-1} region is a dead giveaway for the presence of a carbonyl group*. Within that range, carboxylic acids, esters, ketones, and aldehydes tend to absorb in the shorter wavelength end (1700-1750 cm-1), while conjugated unsaturated ketones and amides tend to absorb on the longer wavelength end (1650-1700 cm^{-1}).

The jagged peak at approximately 2900-3000 cm^{-1} is characteristic of tetrahedral carbon-hydrogen bonds. This peak is not terribly useful, as just about every organic molecule that you will have occasion to analyze has these bonds. Nevertheless, it can serve as a familiar reference point to orient yourself in a spectrum.

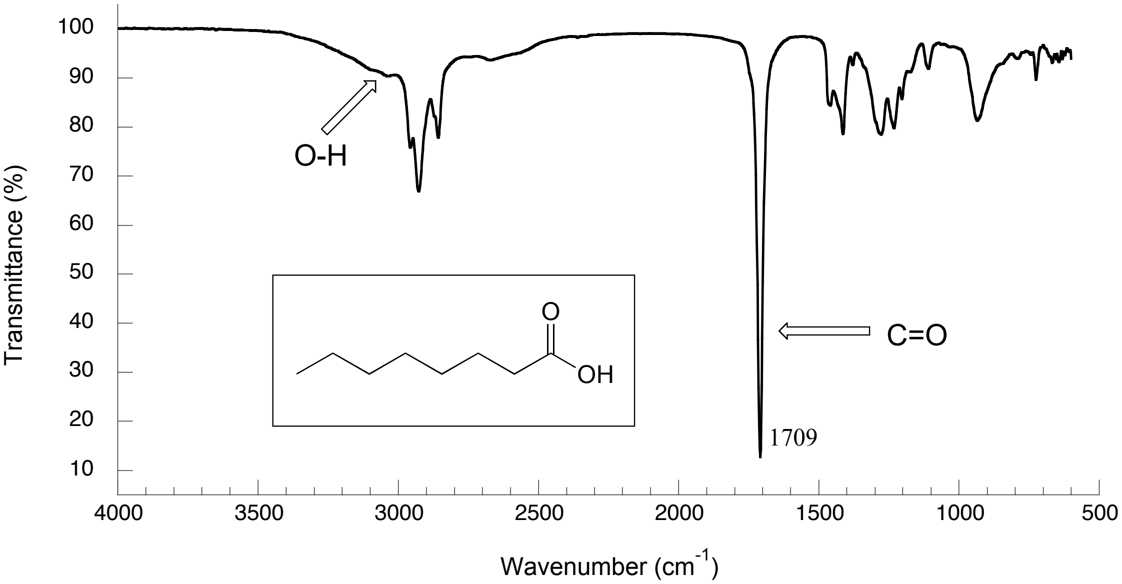
You will notice that there are many additional peaks in this spectrum in the longer-wavelength 400 -1400 cm^{-1} region. This part of the spectrum is called the **fingerprint region**. While it is usually very difficult to pick out any specific functional group identifications from this region, it does, nevertheless, contain valuable information. The reason for this is suggested by the name: just like a human fingerprint, the pattern of absorbance peaks in the fingerprint region is unique to every molecule, meaning that the data from an unknown sample can be compared to the IR spectra of known standards in order to make a positive identification. In the mid-1990's, for example, several paintings were identified as forgeries because scientists were able to identify the IR footprint region of red and yellow pigment compounds that would not have been available to the artist who supposedly created the painting (for more details see Chemical and Engineering News, Sept 10, 2007).

Now, let's take a look at the IR spectrum for 1-hexanol.



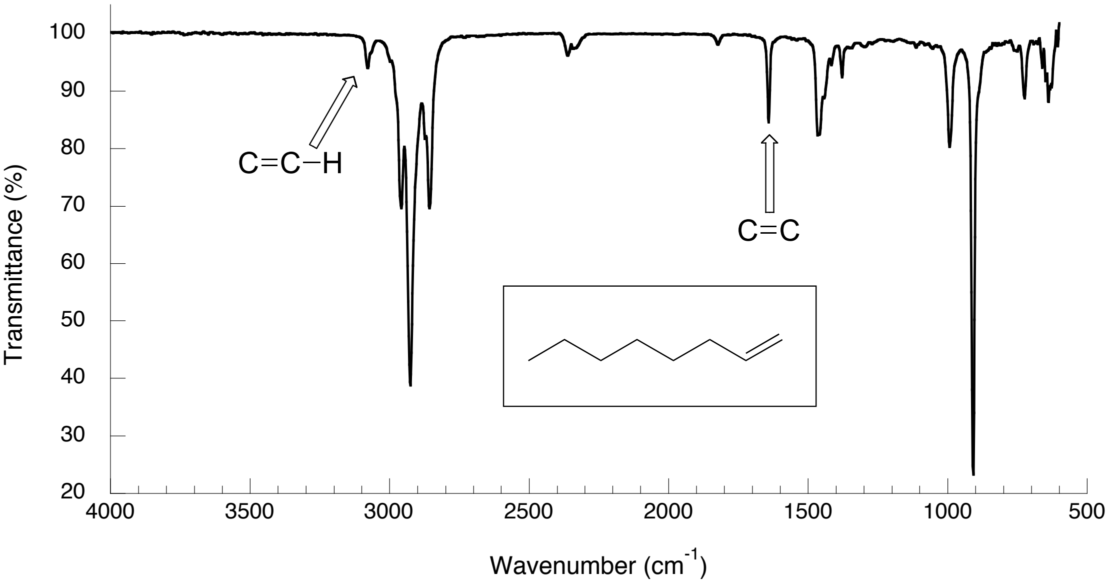
As you can see, the carbonyl peak is gone, and in its place is a very broad 'mountain' centered at about 3400 cm^{-1}. This signal is characteristic of the O-H stretching mode of alcohols, and is a dead giveaway for the presence of an alcohol group. The breadth of this signal is a consequence of hydrogen bonding between molecules.

In the spectrum of octanoic acid we see, as expected, the characteristic carbonyl peak, this time at 1709 cm^{-1}.



We also see a low, broad absorbance band that looks like an alcohol, except that it is displaced slightly to the right (long-wavelength) side of the spectrum, causing it to overlap to some degree with the C-H region. This is the characteristic carboxylic acid O-H single bond stretching absorbance.

The spectrum for 1-octene shows two peaks that are characteristic of alkenes: the one at 1642 cm^{-1} is due to stretching of the carbon-carbon double bond, and the one at 3079 cm-1 is due to stretching of the σ bond between the alkene carbons and their attached hydrogens.



Alkynes have characteristic IR absorbance peaks in the range of 2100-2250 cm^{-1} due to stretching of the carbon-carbon triple bond, and terminal alkenes can be identified by their absorbance at about 3300 cm-1, due to stretching of the bond between the sp-hybridized carbon and the terminal hydrogen.

It is possible to identify other functional groups such as amines and ethers, but the characteristic peaks for these groups are considerably more subtle and/or variable, and often are overlapped with peaks from the fingerprint region. For this reason, we will limit our discussion here to the most easily recognized functional groups, which are summarized in table 1 in the tables section at the end of the text.

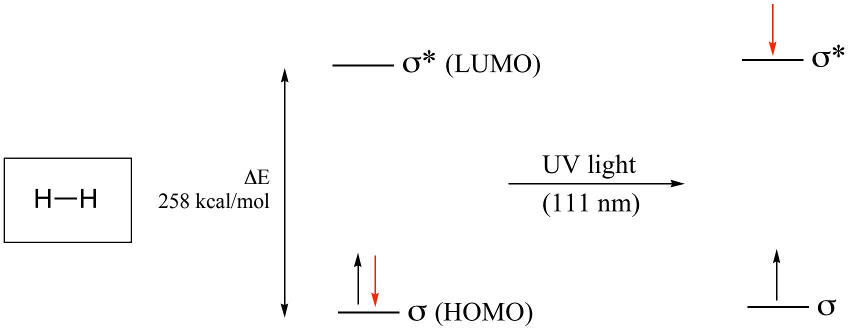
As you can imagine, obtaining an IR spectrum for a compound will not allow us to figure out the complete structure of even a simple molecule, unless we happen to have a reference spectrum for comparison. In conjunction with other analytical methods, however, IR spectroscopy can prove to be a very valuable tool, given the information it provides about the presence or absence of key functional groups. IR can also be a quick and convenient way for a chemist to check to see if a reaction has proceeded as planned. If we were to run a reaction in which we wished to convert cyclohexanone to cyclohexanol, for example, a quick comparison of the IR spectra of starting compound and product would tell us if we had successfully converted the ketone group to an alcohol (this type of reaction is discussed in detail in chapter 16).

Section 4.3: Ultraviolet and visible spectroscopy

4.3A: Electronic transitions

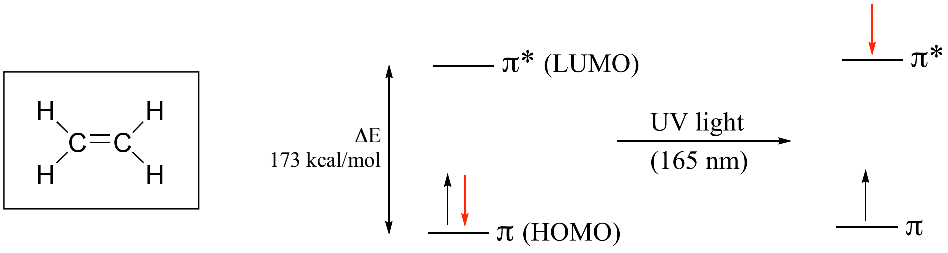
While interaction with infrared light causes molecules to undergo vibrational transitions, the shorter wavelength, higher energy radiation in the UV (200-400 nm) and visible (400-700 nm) range of the electromagnetic spectrum causes many organic molecules to undergo **electronic transitions**. What this means is that when the energy from UV or visible light is absorbed by a molecule, one of its electrons jumps from a lower energy to a higher energy molecular orbital.

Let's take as our first example the simple case of molecular hydrogen, H_2. As you may recall from section 2.1A, the molecular orbital picture for the hydrogen molecule consists of one bonding σ MO, and a higher energy antibonding σ* MO. When the molecule is in the ground state, both electrons are paired in the lower-energy bonding orbital – this is the Highest Occupied Molecular Orbital (HOMO). The antibonding σ* orbital, in turn, is the Lowest Unoccupied Molecular Orbital (LUMO).



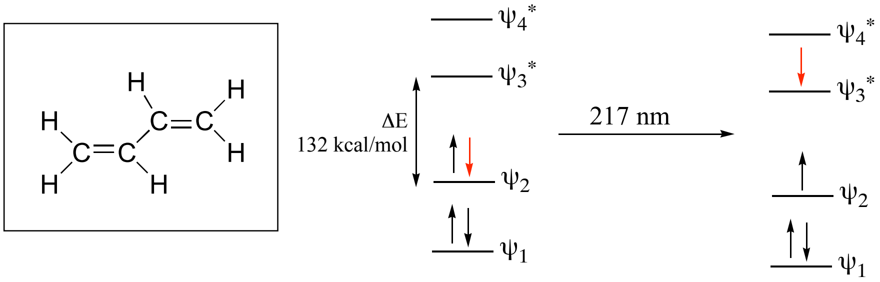
If the molecule is exposed to light of a wavelength with energy equal to ΔE, the HOMO-LUMO energy gap, this wavelength will be absorbed and the energy used to bump one of the electrons from the HOMO to the LUMO – in other words, from the σ to the σ* orbital. This is referred to as a **σ → σ* transition**. ΔE for this electronic transition is 258 kcal/mol, corresponding to light with a wavelength of 111 nm.

When a double-bonded molecule such as ethene (common name ethylene) absorbs light, it undergoes a **π → π* transition.** Because π → π* energy gaps are narrower than σ → σ* gaps, ethene absorbs light at 165 nm - a longer wavelength than molecular hydrogen.



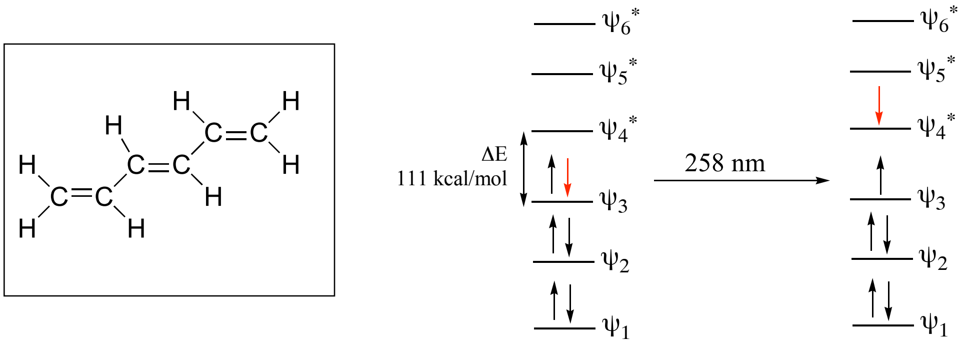
The electronic transitions of both molecular hydrogen and ethene are too energetic to be accurately recorded by standard UV spectrophotometers, which generally have a range of 220 – 700 nm. Where UV-vis spectroscopy becomes useful to most organic and biological chemists is in the study of molecules with conjugated π systems. In these groups, the energy gap for π → π* transitions is smaller than for isolated double bonds, and thus the wavelength absorbed is longer. Molecules or parts of molecules that absorb light strongly in the UV-vis region are called **chromophores**.

Let's revisit the MO picture for 1,3-butadiene, the simplest conjugated system (see section 2.1B). Recall that we can draw a diagram showing the four π MO's that result from combining the four $2p_z$ atomic orbitals. Two of these - ψ_1 and ψ_2 – are bonding, while ψ_3* and ψ_4* are antibonding.

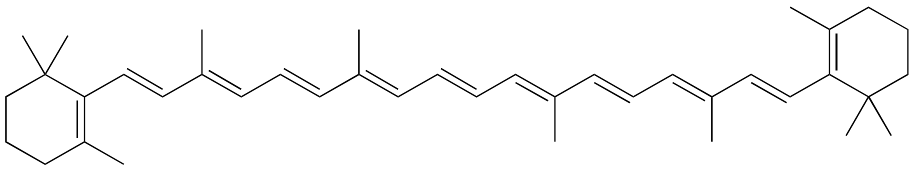


Comparing this MO picture to that of ethene, our isolated π-bond example, we see that the HOMO-LUMO energy gap is indeed smaller for the conjugated system. 1,3-butadiene absorbs UV light with a wavelength of 217 nm.

As conjugated π systems become larger, the energy gap for a π → π* transition becomes increasingly narrow, and the wavelength of light absorbed correspondingly becomes longer. The absorbance due to the π → π* transition in 1,3,5-hexatriene, for example, occurs at 258 nm, corresponding to a ΔE of 111 kcal/mol.

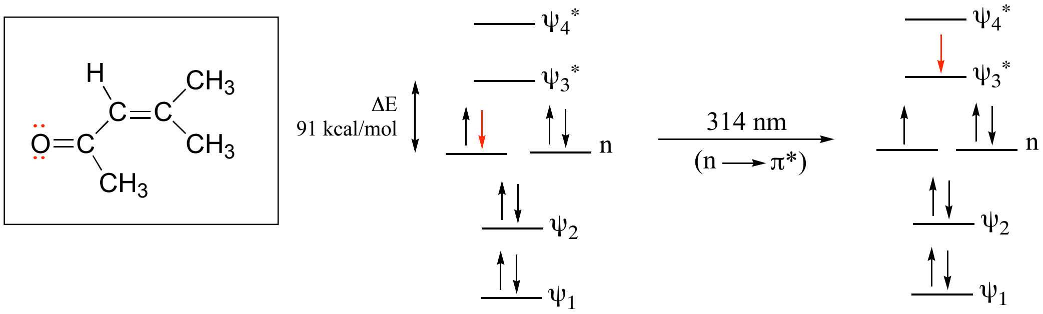


In molecules with extended π systems, the π → π* energy gap becomes so small that absorption occurs in the visible rather then the UV region of the electromagnetic spectrum. β-carotene, with its system of 11 conjugated double bonds, absorbs light with wavelengths in the blue region of the visible spectrum while allowing other visible wavelengths – mainly those in the red-yellow region - to be transmitted. This is why carrots are orange.



β-carotene

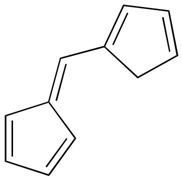
The conjugated π system in 4-methyl-3-penten-2-one gives rise to a strong UV absorbance at $\lambda_{max} = 236$ nm due to a $\pi \rightarrow \pi^*$ transition. However, this molecule also absorbs at $\lambda_{max} = 314$ nm. This second absorbance is due to the transition of a non-bonding (lone pair) electron on the oxygen up to a π^* antibonding MO:



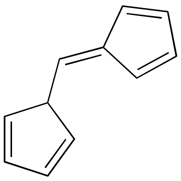
This is referred to as an **n→π* transition**. The nonbonding (n) MO's are higher in energy than the highest bonding π orbitals, so the energy gap for an n→π^* transition is smaller that that of a $\pi \rightarrow \pi^*$ transition – and thus the n→π^* peak is at a longer wavelength. In general, n→π^* transitions are weaker (less light absorbed) than those due to $\pi \rightarrow \pi^*$ transitions.

Exercise 4.3: what is value of ΔE for the $\pi \rightarrow \pi^*$ transition in 4-methyl-3-penten-2-one?

Exercise 4.4: Which of the following molecules would you expect absorb at a longer wavelength in the UV region of the electromagnetic spectrum? Explain your answer.



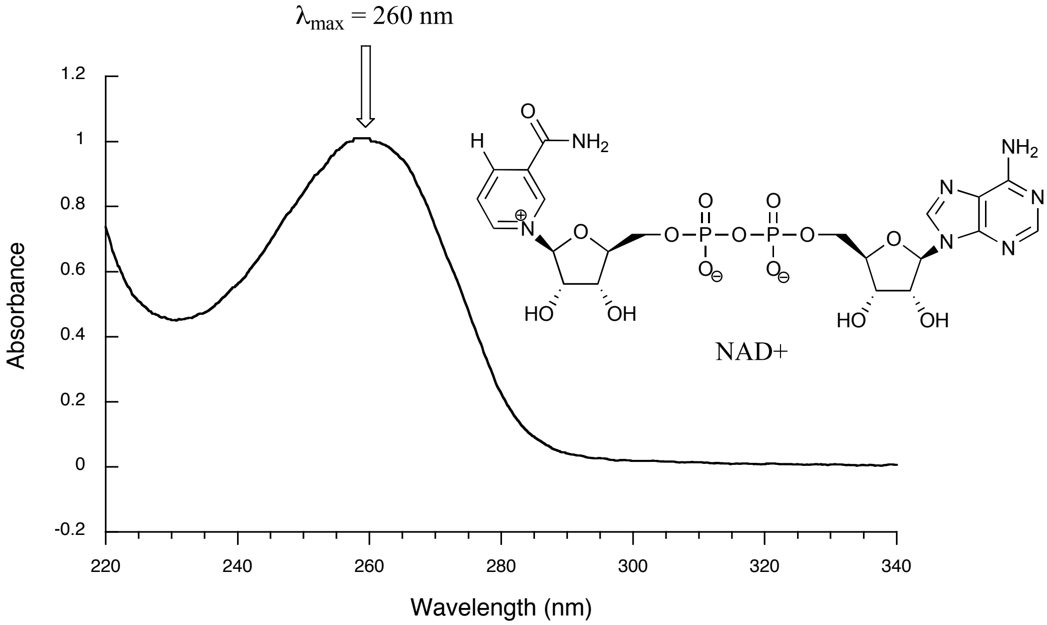
A



B

4.3B: Looking at UV-vis spectra

We have been talking in general terms about how molecules absorb UV and visible light – now let's look at some actual examples of data from a UV-vis absorbance spectrophotometer. The basic setup is the same as for IR spectroscopy: radiation with a range of wavelengths is directed through a sample of interest, and a detector records which wavelengths were absorbed and to what extent the absorption occurred. Below is the absorbance spectrum of an important biological molecule called nicotinamide adenine dinucleotide, abbreviated NAD$^+$. This compound absorbs light in the UV range due to the presence of conjugated π-bonding systems.



You'll notice that this UV spectrum is much simpler than the IR spectra we saw earlier: this one has only one peak, although many molecules have more than one. Notice also that the convention in UV-vis spectroscopy is to show the baseline at the bottom of the graph with the peaks pointing up. Wavelength values on the x-axis are generally measured in nanometers (nm) rather than in cm^{-1} as is the convention in IR spectroscopy.

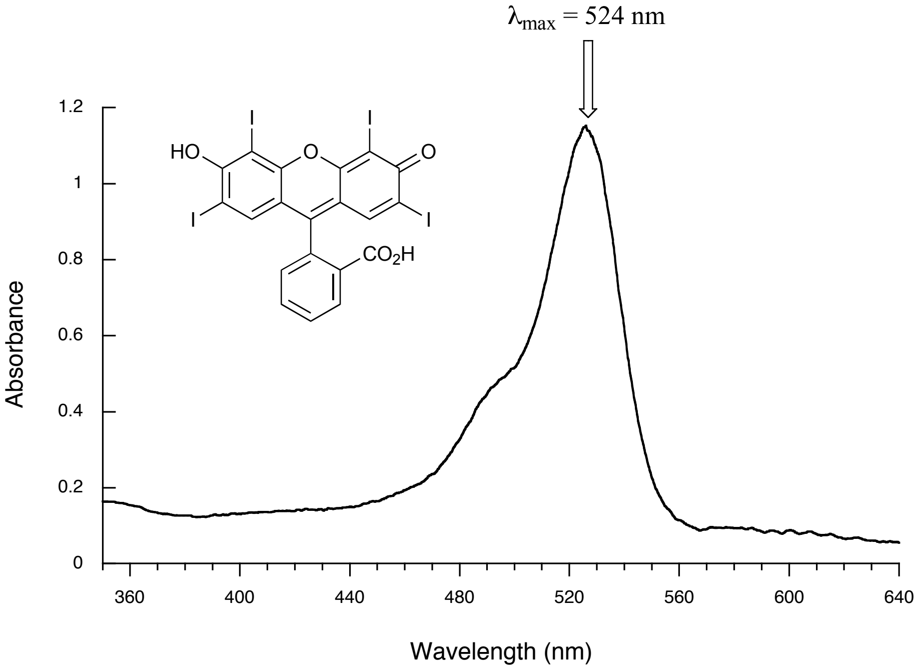
Peaks in UV spectra tend to be quite broad, often spanning well over 20 nm at half-maximal height. Typically, there are two things that we look for and record from a UV-Vis spectrum.. The first is $\lambda_{\text{max,}}$ which is the wavelength at maximal light absorbance. As you can see, NAD$^+$ has $\lambda_{\text{max,}}$ = 260 nm. We also want to record how much light is absorbed at $\lambda_{\text{max.}}$ Here we use a unitless number called **absorbance**, abbreviated 'A'. This contains the same information as the 'percent transmittance' number used in IR spectroscopy, just expressed in slightly different terms. To calculate absorbance at a given wavelength, the computer in the spectrophotometer simply takes the intensity of light at that wavelength *before* it passes through the sample (I_0), divides this value by the intensity of the same wavelength *after* it passes through the sample (I), then takes the $\log_{10}$ of that number:

equation 4.3 $$A = \log I_0/I$$

You can see that the absorbance value at 260 nm (A_{260}) is about 1.0 in this spectrum.

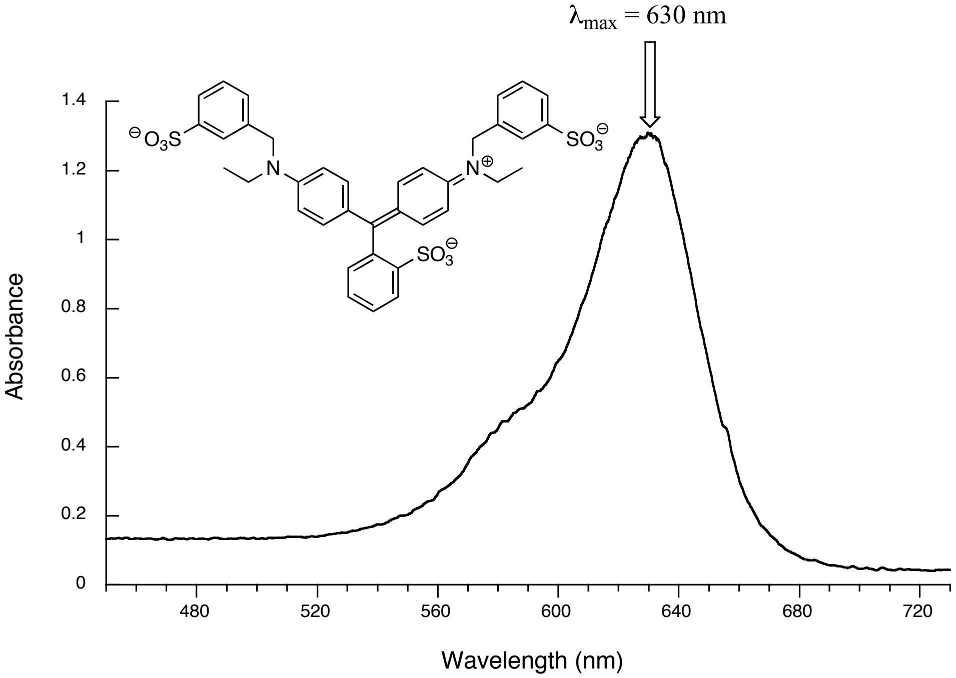
Exercise 4.5: Express A = 1.0 in terms of percent transmittance (%T, the unit usually used in IR spectroscopy (and sometimes in UV-vis as well).

Here is the absorbance spectrum of the common food coloring Red #3:



Here, we see that the extended system of conjugated π bonds causes the molecule to absorb light in the visible range. Because the λ_{max} of 524 nm falls within the green region of the spectrum, the compound appears red to our eyes.

Now, take a look at the spectrum of another food coloring, Blue #1:



Here, maximum absorbance is at 630 nm, in the orange range of the visible spectrum, and the compound appears blue.

4.3C: Applications of UV spectroscopy in organic and biological chemistry

UV-vis spectroscopy has many different applications in organic and biological chemistry. One of the most basic of these applications is the use of the **Beer - Lambert Law** to determine the concentration of a chromophore. You most likely have performed a Beer – Lambert experiment in a previous chemistry lab. The law is simply an application of the observation that, within certain ranges, the absorbance of a chromophore at a given wavelength varies in a linear fashion with its concentration: the higher the concentration of the molecule, the greater its absorbance. If we divide the observed value of A at λ_{max} by the concentration of the sample (c, in mol/L), we obtain the **molar absorptivity**, or **extinction coefficient** (ε), which is a characteristic value for a given compound.

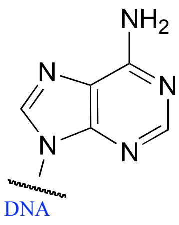
equation 4.4 $\varepsilon = A/c$

The absorbance will also depend, of course, on the **path length** - in other words, the distance that the beam of light travels though the sample. In most cases, sample holders are designed so that the path length is equal to 1 cm, so the units for molar absorptivity are mol·L^{-1}cm^{-1}.

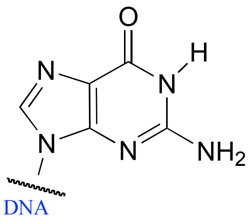
If we look up the value of ε for our compound at λ_{max}, and we measure absorbance at this wavelength, we can easily calculate the concentration of our sample. As an example, for NAD$^+$ the literature value of ε at 260 nm is 18,000 mol·L^{-1}cm^{-1}. In our NAD$^+$ spectrum we observed $A_{260} = 1.0$, so using equation 4.4 and solving for concentration we find that our sample is 5.6×10^{-5} M.

Exercise 4.6: The literature value of ε for 1,3-pentadiene in hexane is 26,000 mol·L^{-1}cm^{-1} at its λ_{max} of 224 nm. You prepare a sample and take a UV spectrum, finding that $A_{224} = 0.850$. What is the concentration of your sample?

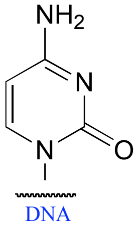
The bases of DNA and RNA are good chromophores:



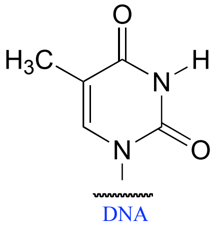
Adenine



Guanine



Cytosine

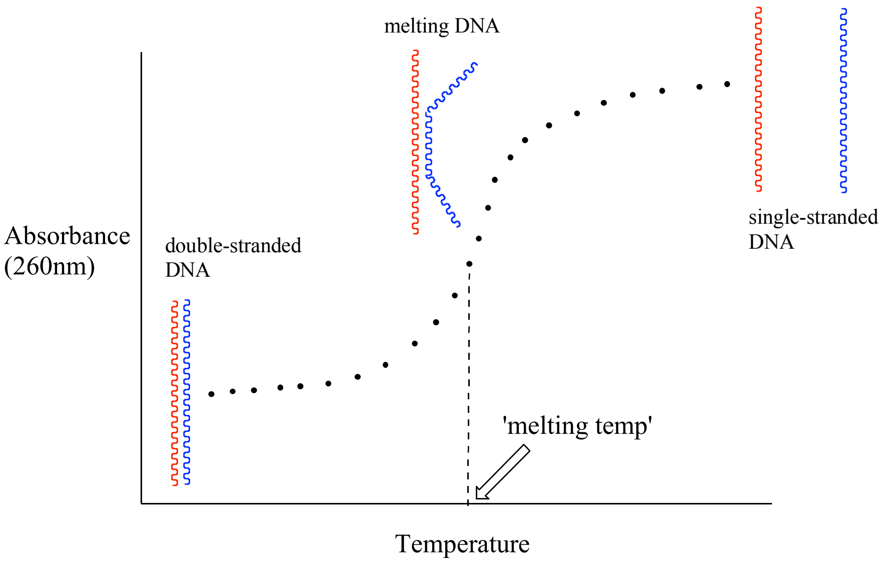


Thymine

Biochemists and molecular biologists often determine the concentration of a DNA sample by assuming an average value of $\varepsilon = 0.020$ ng^{-1}·μL for double-stranded DNA at its λ_{max} of 260 nm (notice that concentration in this application is expressed in mass/volume rather than molarity: ng/μL is often a convenient unit for DNA concentration when doing molecular biology).

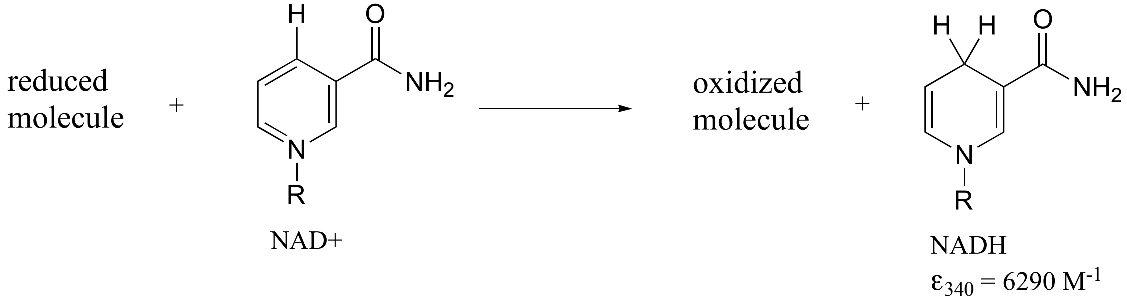
Exercise 4.7: 50 μL of an aqueous sample of double stranded DNA is dissolved in 950 μL of water. This diluted solution has a maximal absorbance of 0.326 at 260 nm. What is the concentration of the original (more concentrated) DNA sample, expressed in μg/μL?

Because the extinction coefficient of double stranded DNA is slightly lower than that of single stranded DNA, we can use UV spectroscopy to monitor a process known as DNA melting. If a short stretch of double stranded DNA is gradually heated up, it will begin to 'melt', or break apart, as the temperature increases (recall that two strands of DNA are held together by a specific pattern of hydrogen bonds formed by 'base-pairing').

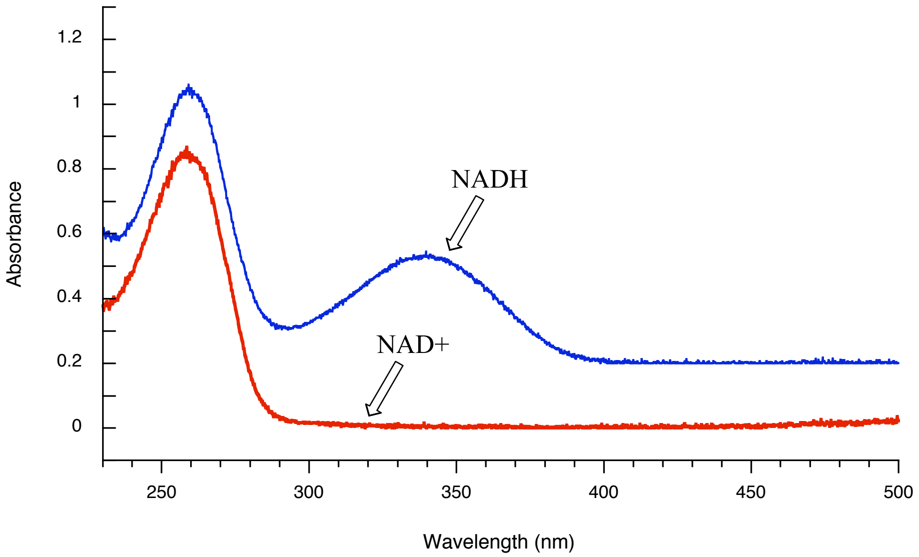


As melting proceeds, the absorbance value for the sample increases, eventually reaching a high plateau as all of the double-stranded DNA breaks apart, or 'melts'. The mid-point of this process, called the 'melting temperature', provides a good indication of how tightly the two strands of DNA are able to bind to each other.

In section 16.8 we will see how the Beer - Lambert Law and UV spectroscopy provides us with a convenient way to follow the progress of many different enzymatic redox (oxidation-reduction) reactions. In biochemistry, oxidation of an organic molecule often occurs concurrently with reduction of nicotinamide adenine dinucleotide (NAD^+, the compound whose spectrum we saw earlier in this section) to NADH:

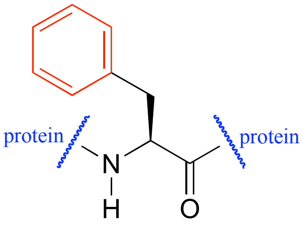


Both NAD^+ and NADH absorb at 260 nm. However NADH, unlike NAD^+, has a second absorbance band with $\lambda_{max} = 340$ nm and $\varepsilon = 6290\ mol \cdot L^{-1}cm^{-1}$. The figure below shows the spectra of both compounds superimposed, with the NADH spectrum offset slightly on the y-axis:

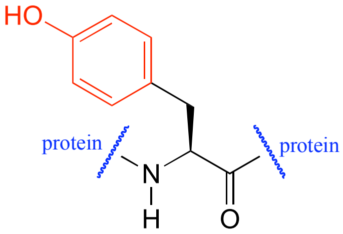


By monitoring the absorbance of a reaction mixture at 340 nm, we can 'watch' NADH being formed as the reaction proceeds, and calculate the rate of the reaction.

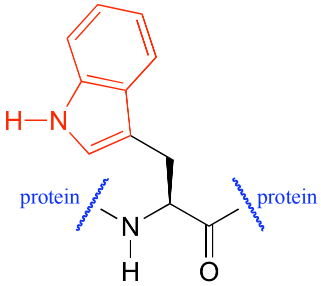
UV spectroscopy is also very useful in the study of proteins. Proteins absorb light in the UV range due to the presence of the aromatic amino acids tryptophan, phenylalanine, and tyrosine, all of which are chromophores.



phenylalanine



tyrosine



tryptophan

Biochemists frequently use UV spectroscopy to study conformational changes in proteins - how they change shape in response to different conditions. When a protein undergoes a conformational shift (partial unfolding, for example), the resulting change in the environment around an aromatic amino acid chromophore can cause its UV spectrum to be altered.

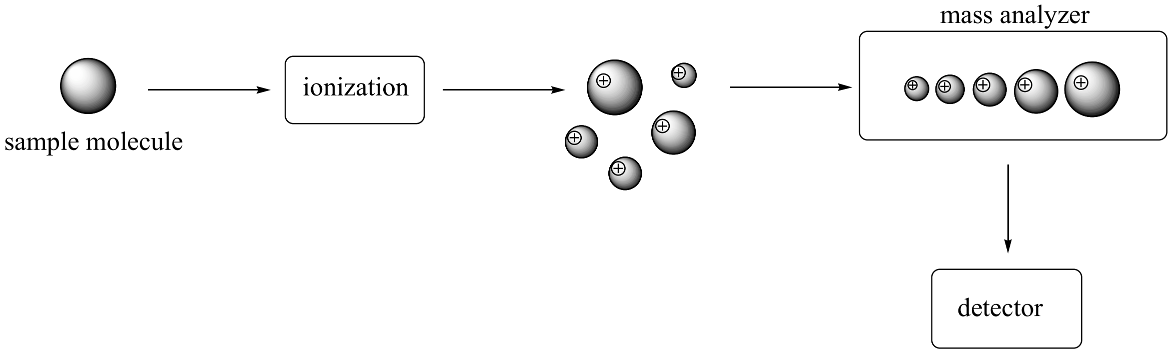
Section 4.4: Mass Spectrometry

Our third and final analytical technique for discussion in this chapter does *not* fall under the definition of spectroscopy, as it does not involve the absorbance of light by a molecule. In mass spectrometry (MS), we are interested in the mass - and therefore the molecular weight - of our compound of interest, and often the mass of fragments that are produced when the molecule is caused to break apart.

4.4A: The basics of a mass spectrometry

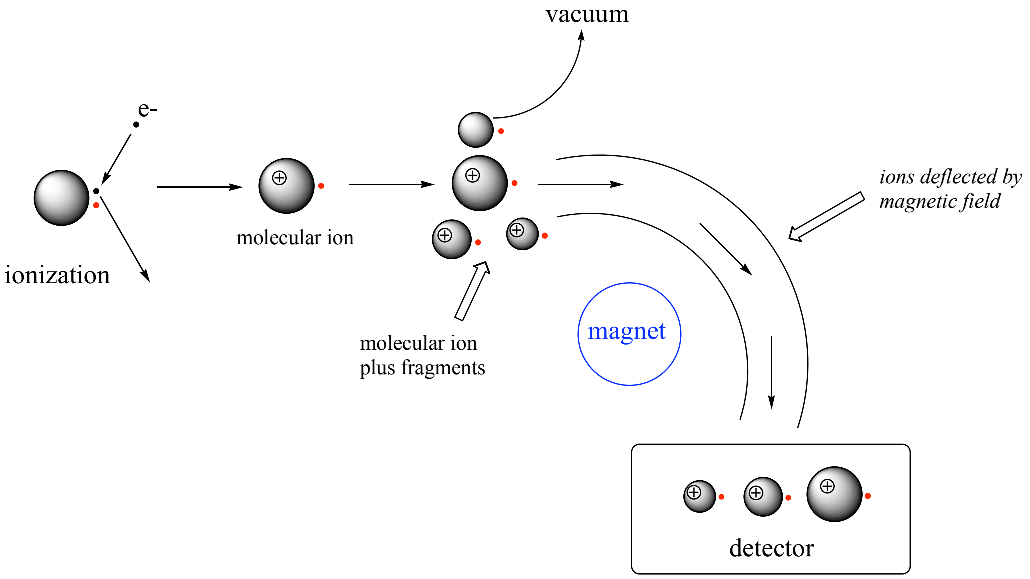
There are many different types of MS instruments, but they all have the same three essential components. First, there is an ionization source, where the molecule is given a positive electrical charge, either by removing an electron or by adding a proton. Depending on the ionization method used, the ionized molecule may or may not break apart into a population of smaller fragments. In the figure below, some of the sample molecules remain whole, while others fragment into smaller pieces.

Next in line there is a mass analyzer, where the cationic fragments are separated according to their mass.



Finally, there is a detector, which detects and quantifies the separated ions.

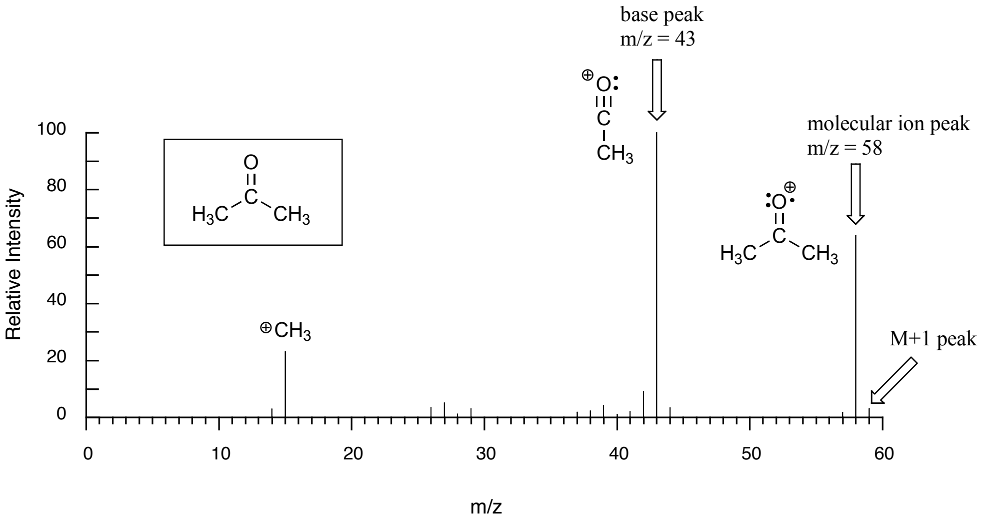
One of the more common types of MS techniques used in the organic laboratory is **electron ionization**. In the ionization source, the sample molecule is bombarded by a high-energy electron beam, which has the effect of knocking a valence electron off of the molecule to form a **radical cation**. Because a great deal of energy is transferred by this bombardment process, the radical cation quickly begins to break up into smaller fragments, some of which are positively charged and some of which are neutral. The neutral fragments are either adsorbed onto the walls of the chamber or are removed by a vacuum source. In the mass analyzer component, the positively charged fragments and any remaining unfragmented **molecular ions** are accelerated down a tube by an electric field.


vacuum
e-
ionization
molecular ion
molecular ion plus fragments
ions deflected by magnetic field
magnet
detector

This tube is curved, and the ions are deflected by a strong magnetic field. Ions of different mass to charge (m/z) ratios are deflected to a different extent, resulting in a 'sorting' of ions by mass (virtually all ions have charges of z = +1, so sorting by the mass to charge ratio is the same thing as sorting by mass). A detector at the end of the curved flight tube records and quantifies the sorted ions.

4.4B: Looking at mass spectra

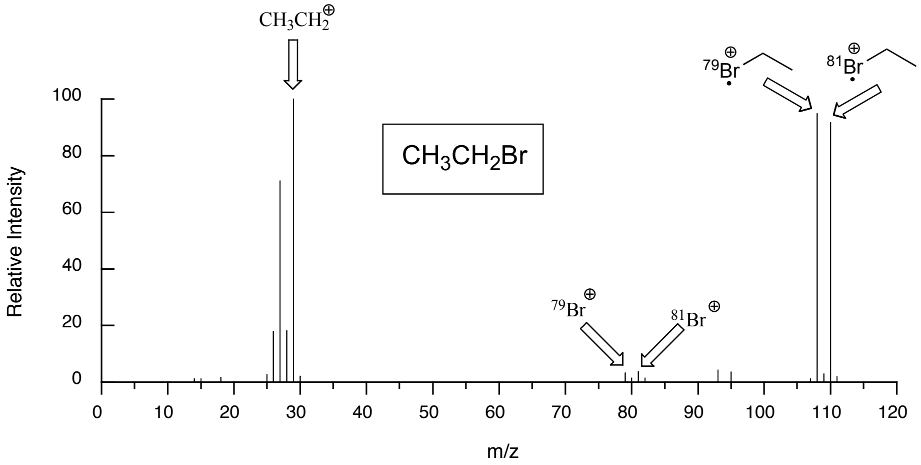
Below is typical output for an electron-ionization MS experiment (MS data in the section is derived from the Spectral Database for Organic Compounds, a free, web-based service provided by AIST in Japan: http://riodb01.ibase.aist.go.jp/sdbs/).



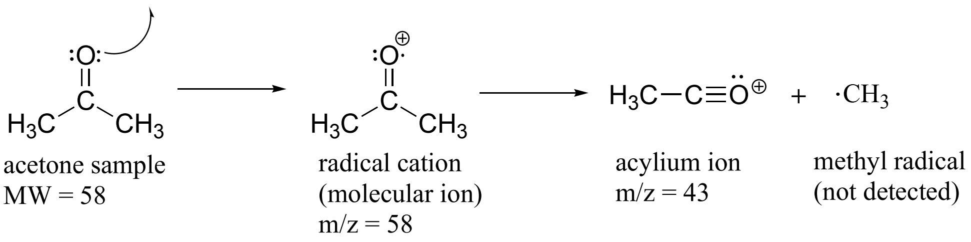
The sample is acetone. On the horizontal axis is the value for m/z (as we stated above, the charge z is almost always +1, so in practice this is the same as mass). On the vertical axis is the relative abundance of each ion detected. On this scale, the most abundant ion, called the

base peak, is set to 100%, and all other peaks are recorded relative to this value. For acetone, the base peak is at m/z = 43 - we will discuss the formation of this fragment a bit later. The molecular weight of acetone is 58, so we can identify the peak at m/z = 58 as that corresponding to the **molecular ion peak**, or **parent peak**. Notice that there is a small peak at m/z = 59: this is referred to as the **M+1 peak**. How can there be an ion that has a greater mass than the molecular ion? Simple: a small fraction - about 1.1% - of all carbon atoms in nature are actually the ^{13}C rather than the ^{12}C isotope. The ^{13}C isotope is, of course, heavier than ^{12}C by 1 mass unit. In addition, about 0.015% of all hydrogen atoms are actually deuterium, the ^{2}H isotope. So the M+1 peak represents those few acetone molecules in the sample which contained either a ^{13}C or ^{2}H.

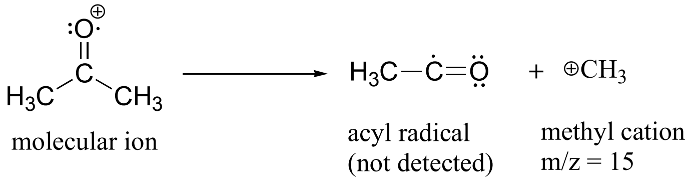
Molecules with lots of oxygen atoms sometimes show a small **M+2 peak** (2 m/z units greater than the parent peak) in their mass spectra, due to the presence of a small amount of ^{18}O (the most abundant isotope of oxygen is ^{16}O). Because there are two abundant isotopes of both chlorine (about 75% ^{35}Cl and 25% ^{37}Cl) and bromine (about 50% ^{79}Br and 50% ^{81}Br), chlorinated and brominated compounds have very large and recognizable M+2 peaks. Fragments containing both isotopes of Br can be seen in the mass spectrum of ethyl bromide:



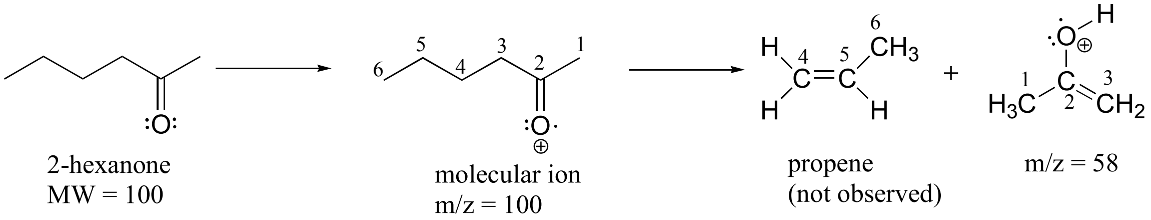
Much of the utility in electron-ionization MS comes from the fact that the radical cations generated in the electron-bombardment process tend to fragment in predictable ways. Detailed analysis of the typical fragmentation patterns of different functional groups is beyond the scope of this text, but it is worthwhile to see a few representative examples, even if we don't attempt to understand the exact process by which the fragmentation occurs. We saw, for example, that the base peak in the mass spectrum of acetone is m/z = 43. This is the result of cleavage at the 'alpha' position - in other words, at the carbon-carbon bond adjacent to the carbonyl. Alpha cleavage results in the formation of an acylium ion (which accounts for the base peak at m/z = 43) and a methyl radical, which is neutral and therefore not detected.



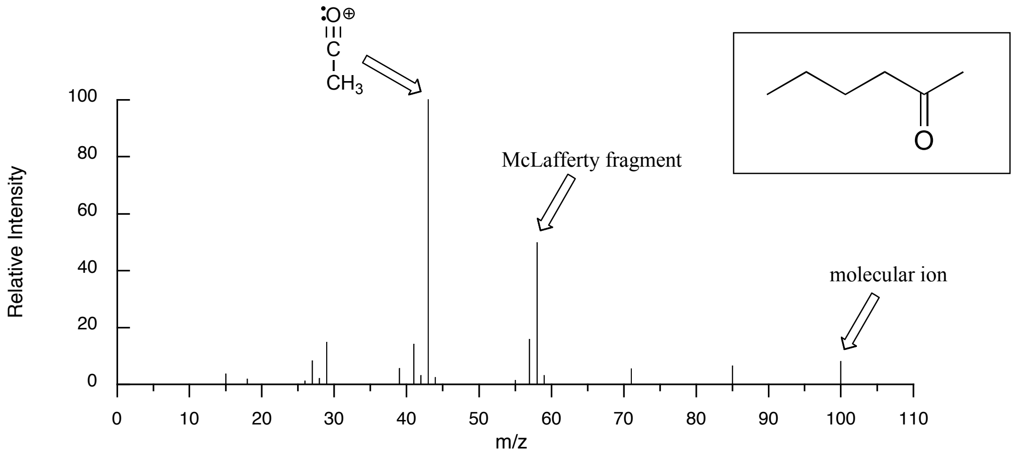
After the parent peak and the base peak, the next largest peak, at a relative abundance of 23%, is at m/z = 15. This, as you might expect, is the result of formation of a methyl cation, in addition to an acyl radical (which is neutral and not detected).



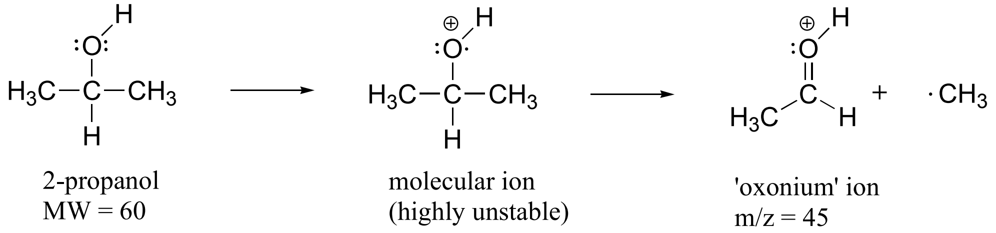
A common fragmentation pattern for larger carbonyl compounds is called the McLafferty rearrangement'.

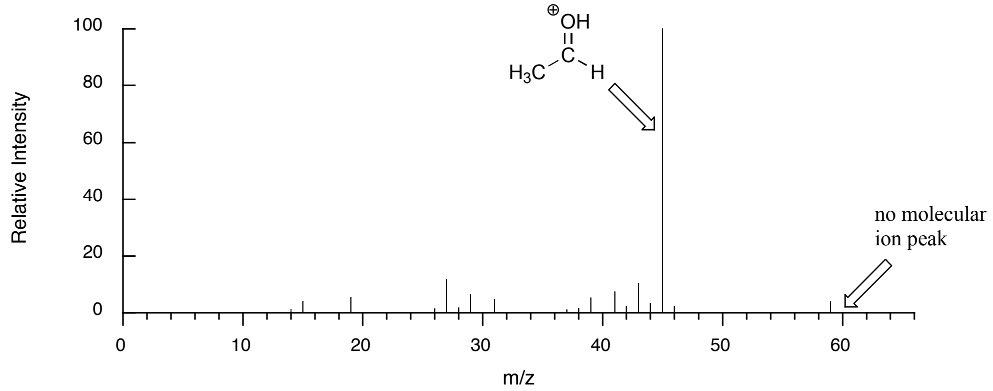


The mass spectrum of 2-hexanone shows a 'McLafferty fragment' at m/z = 58, while the propene fragment is not observed because it is a neutral species (remember, only cationic fragments are observed in MS). The base peak in this spectrum is again an acylium ion.



When alcohols are subjected to electron ionization MS, the molecular ion is highly unstable and thus a parent peak is often not detected. Often the base peak is from an 'oxonium' ion.





Other functional groups have predictable fragmentation patterns as well. By carefully analyzing the fragmentation information that a mass spectrum provides, a knowledgeable spectrometrist can often 'put the puzzle together' and make some very confident predictions about the structure of the starting sample.

Exercise 4.8: The mass spectrum of an aldehyde gives prominent peaks at *m/z* = 59 (12%, highest value of *m/z* in the spectrum), 58 (85%), and 29 (100%), as well as others. Propose a structure, and identify the three species whose *m/z* values were listed.

4.4C: Gas Chromatography - Mass Spectrometry

Quite often, mass spectrometry is used in conjunction with a separation technique called gas chromatography (GC). The combined GC-MS procedure is very useful when dealing with a sample that is a mixture of two or more different compounds, because the various compounds are separated from one another before being subjected individually to MS analysis. We will not go into the details of gas chromatography here, although if you are taking an organic laboratory course you might well get a chance to try your hand at GC, and you will almost certainly be exposed to the conceptually analogous techniques of thin layer and column chromatography. Suffice it to say that in GC, a very small amount of a liquid sample is vaporized, injected into a long, coiled metal column, and pushed though the column by helium gas. Along the way, different compounds in the sample stick to the walls of the column to different extents, and thus travel at different speeds and emerge separately from the end of the column. In GC-MS, each purified compound is sent directly from the end of GC column into the MS instrument, so in the end we get a separate mass spectrum for each of the compounds in the original mixed sample. Because a compound's MS spectrum is a very reliable and reproducible 'fingerprint', we can instruct the instrument to search an MS database and identify each compound in the sample.

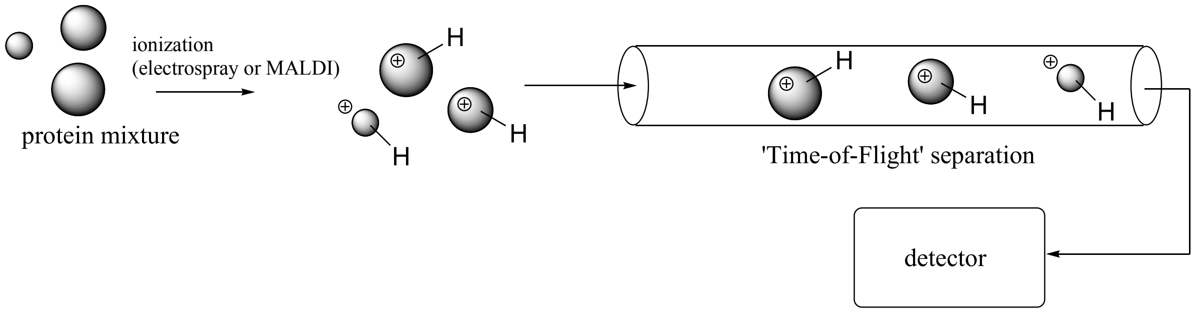
The extremely high sensitivity of modern GC-MS instrumentation makes it possible to detect and identify very small trace amounts of organic compounds. GC-MS is being used increasingly by environmental chemists to detect the presence of harmful organic contaminants in food and water samples. Airport security screeners also use high-speed GC-MS instruments to look for residue from bomb-making chemicals on checked luggage.

4.4D: Mass spectrometry of proteins - applications in proteomics

Mass spectrometry has become in recent years an increasingly important tool in the field of **proteomics**. Traditionally, protein biochemists tend to study the structure and function of individual proteins. Proteomics researchers, in contrast, want to learn more about how large numbers of proteins in a living system interact with each other, and how they respond to changes in the state of the organism. One very important subfield of proteomics is the search for protein **biomarkers** for human disease. These can be proteins which are present in greater quantities in a sick person than in a healthy person, and their detection and identification can provide medical researchers with valuable information about possible causes or treatments. Detection in a healthy person of a known biomarker for a disease such as diabetes or cancer could also provide doctors with an early warning that the patient may be especially susceptible, so that preventive measures could be taken to prevent or delay onset of the disease.

New developments in MS technology have made it easier to detect and identify proteins that are present in very small quantities in biological samples. Mass spectrometrists who study proteins often use instrumentation that is somewhat different from the electron-ionization, magnetic deflection system described earlier. When proteins are being analyzed, the object is often to ionize the proteins *without* causing fragmentation, so 'softer' ionization methods are required. In one such method, called **electrospray ionization**, the protein sample, in solution, is sprayed into a tube and the molecules are induced by an electric field to pick up extra protons from the solvent. Another common 'soft ionization' method is 'matrix-assisted laser desorption ionization' (**MALDI**). Here, the protein sample is adsorbed onto a solid matrix, and protonation is achieved with a laser.

Typically, both electrospray ionization and MALDI are used in conjunction with a time-of-flight (TOF) mass analyzer component.



The ionized proteins are accelerated by an electrode through a column, and separation is achieved because lighter ions travel at greater velocity than heavier ions with the same overall charge. In this way, the many proteins in a complex biological sample (such as blood plasma, urine, etc.) can be separated and their individual masses determined very accurately. Modern protein MS is extremely sensitive – very recently, scientists were even able to obtain a mass spectrum of *Tyrannosaurus rex* protein from fossilized bone! (Science **2007**, 316, 277).

In one recent study, MALDI-TOF mass spectrometry was used to compare fluid samples from lung transplant recipients who had suffered from tissue rejection to samples from recipients who had not suffered rejection. Three peptides (short proteins) were found to be present at elevated levels specifically in the tissue rejection samples. It is hoped that these peptides might serve as biomarkers to identify patients who are at increased risk of rejecting their transplanted lungs. (Proteomics **2005**, 5, 1705).

Chapter 4 problems

P4.1: Which represents a higher energy frequency of electromagnetic radiation, 1690 cm^{-1} or 3400 cm^{-1}? Express each of these in terms of wavelength (meters) and frequency (Hz).

P4.2: Calculate the value, in kcal/mol, of the ΔE associated with a typical ketone carbonyl stretching transition at 1720 cm^{-1}.

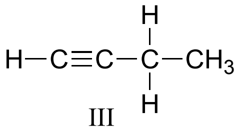
P4.3: Explain how you could use IR spectroscopy to distinguish between compounds I, II, and III.

$H_3C-C\equiv C-CH_3$

I

$H_3C-C\equiv C-CF_3$

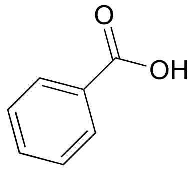
II



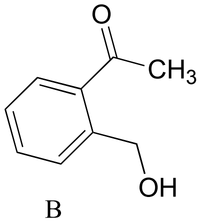
$H-C\equiv C-\overset{H}{\underset{H}{C}}-CH_3$

III

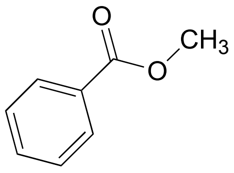
P4.4: Explain how you could use IR spectroscopy to distinguish between compounds A, B, and C.



A

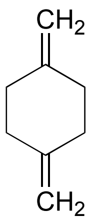


B

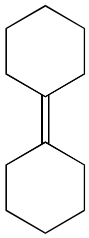


C

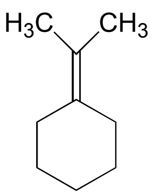
P4.5: Explain how you could use IR spectroscopy to distinguish between compounds X, Y, and Z.



X



Y

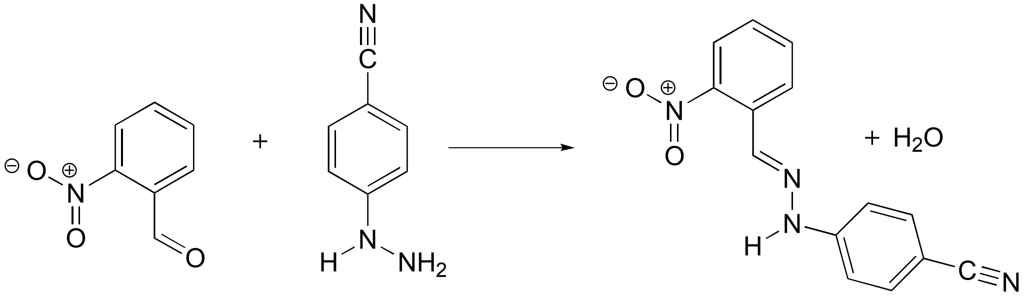


Z

P4.6: A 0.725 mL aqueous solution of NADH shows an absorbance at 340 nm of A = 0.257. Express (in nanamole (nm) units) how much NADH is contained in the sample.

P4.7: A 1 mL enzymatic reaction mixture contains NADH as one of the reactants, and has a starting $A_{340} = 0.345$. After the reaction is allowed to run for five minutes, the researcher records a value of $A_{340} = 0.125$. How many nm of NADH are used up per minute, on average, in this reaction?

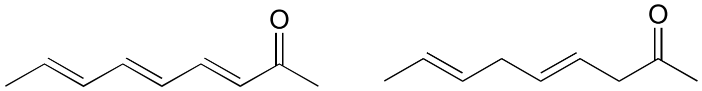
P4.8: The reaction below is of a type that we will study in Chapter 11. While the two starting materials are only slightly colored, the product is an intense orange-red. Account for this observation.



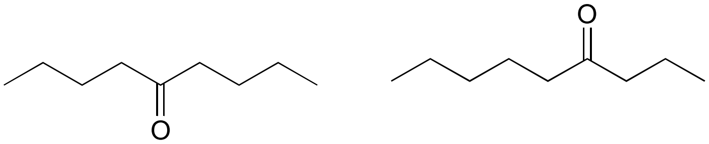
P4.9: Predict five fragments (in addition to the molecular ion) that you would expect to be evident in the mass spectrum of 3-methyl-2-pentanone.

P4.10: One would expect the mass spectrum of cyclohexanone to show a molecular ion peak at m/z = 98. However, the m/z = 98 peak in the cyclohexanone spectrum is unusually tall, compared to the molecular ion peaks in the mass spectra of other ketones such as 2-hexanone or 3-hexanone. Explain.

P4.11: Which would be more useful in distinguishing the two compounds shown below: IR or UV spectroscopy? Explain.



P4.12: Which analytical technique – IR, UV, or MS - could best be used to distinguish between the two compounds below? Explain.



Chapter 5

Structure determination, part II:

Nuclear magnetic resonance spectroscopy

Introduction

In the previous chapter, we learned about three important analytical techniques which allow us to deduce information about the structure of an organic molecule. In IR spectroscopy, transitions in the vibrational states of covalent bonds lead to the absorbance of specific infrared frequencies, telling us about the presence or absence of functional groups in the molecule of interest. In UV-Vis spectroscopy, transitions in the energy levels of electrons in π bonds lead to the absorbance of ultraviolet and visible light, providing us with information about the arrangement of double bonds in a molecule. And in mass spectrometry, we are usually able to learn the molecular weight of a sample molecule, in addition to other kinds of information from analysis of the masses of molecular fragments.

Although all three of these techniques provide us with valuable information about a molecule of interest, in most cases they do not – even in combination – tell us what we, as organic chemists, most want to know about our molecule. Specifically, these techniques do not allow us to determine its overall molecular structure – the framework, in other words, of its carbon-carbon and carbon-hydrogen bonds. It is this information that we need to have in order to be able to draw a Lewis structure of our molecule, and it is this information that is provided by an immensely powerful analytical technique called nuclear magnetic resonance (NMR) spectroscopy.

In NMR, the nuclei of hydrogen, carbon, and other important elements undergo transitions in their magnetic states, leading to the absorbance of radiation in the radio frequency range of the electromagnetic spectrum. By analyzing the signals from these transitions, we learn about the chemical environment that each atom inhabits, including information about the presence of neighboring atoms. In this chapter, we will see how information from NMR, especially when combined with data from IR, UV-Vis, and MS experiments, can make it possible for us to form a complete picture of the atom-to-atom framework of an organic molecule.

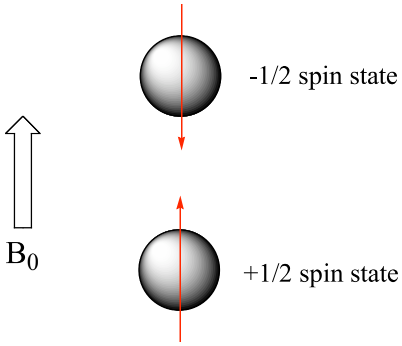
Section 5.1: The origin of the NMR signal

5.1A: NMR-active nuclei

The basis for nuclear magnetic resonance is the observation that many atomic nuclei spin about an axis and generate their own magnetic field, or **magnetic moment**. For reasons that are outside the scope of this text, only those nuclei with an odd number of protons and/or neutrons have a magnetic moment. Fortunately for chemists, several common nuclei, including hydrogen (^1H), the ^{13}C isotope of carbon, the ^{19}F isotope of fluorine, and the ^{31}P isotope of phosphorus, all have magnetic moments and therefore can be observed by NMR – they are, in other words, NMR-active. Other nuclei - such as the common ^{12}C and ^{16}O isotopes of carbon and oxygen - do not have magnetic moments, and are essentially invisible in NMR. Other nuclei such as deuterium (^2H) and nitrogen (^{14}N) have magnetic moments and are NMR-active, but the nature of their magnetic moments is such that these nuclei are more difficult to analyze by NMR. In practice it is ^1H, ^{13}C, ^{19}F, and ^{31}P that are most often observed by NMR spectroscopy. In this chapter, we will develop our understanding of the principles behind NMR spectroscopy by focusing our attention first on the detection of protons in ^1H-NMR experiments (in discussion about NMR, the terms 'hydrogen' and 'proton' are used interchangeably). Much of what we learn, however, will also apply to the detection and analysis of other NMR-active nuclei, and later in the chapter we will shift our attention to NMR experiments involving ^{13}C and ^{31}P atoms.

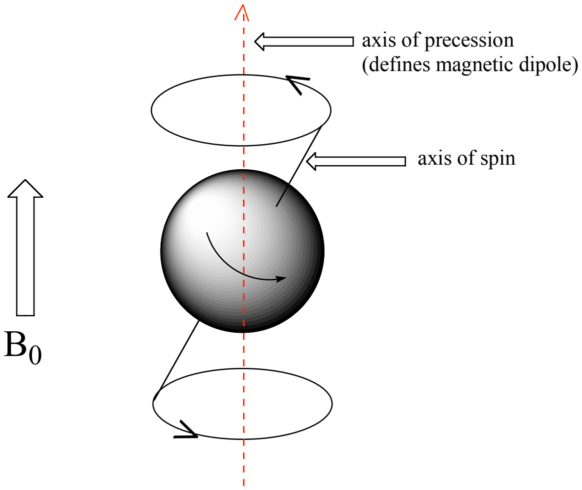
5.1B: Nuclear precession, spin states, and the resonance condition

When a sample of an organic compound is sitting in a flask on a laboratory benchtop, the magnetic moments of its hydrogen atoms are randomly oriented. When the same sample is placed within the field of a very strong magnet in an NMR instrument (this field is referred to by NMR spectroscopists as the **applied field**, abbreviated $\mathbf{B_0}$) each hydrogen will assume one of two possible **spin states**. In what is referred to as the +½ spin state, the hydrogen's magnetic moment is aligned *with* the direction of B_0, while in the -½ spin state it is aligned *opposed to* the direction of B_0.



Because the +½ spin state is slightly lower in energy, in a large population of organic molecules slightly more than half of the hydrogen atoms will occupy this state, while slightly less than half will occupy the –½ state. *The value of ΔE, the difference in energy between the two spin states, increases with increasing strength of B_0.* This last statement is in italics because it is one of the key ideas in NMR spectroscopy, as we shall soon see.

At this point, we need to look a little more closely at how a proton spins in an applied magnetic field. You may recall playing with spinning tops as a child. When a top slows down a little and the spin axis is no longer completely vertical, it begins to exhibit **precessional motion**, as the spin axis rotates slowly around the vertical. In the same way, hydrogen atoms spinning in an applied magnetic field also exhibit precessional motion about a vertical axis. It is this axis (which is either parallel or antiparallel to B_0) that defines the proton's magnetic moment. In the figure below, the proton is in the +1/2 spin state.

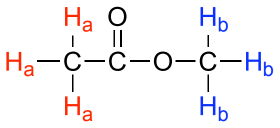


The **frequency of precession** (also called the **Larmour frequency,** abbreviated ν_L) is simply the number of times per second that the proton precesses in a complete circle. A proton`s precessional frequency increases with the strength of B_0.

If a proton that is precessing in an applied magnetic field is exposed to electromagnetic radiation of a frequency ν that matches its precessional frequency ν_L, we have a condition called **resonance**. *In the resonance condition, a proton in the lower-energy +½ spin* *state (aligned with B_0) will transition (flip) to the higher energy –½ spin state (opposed to B_0). In doing so, it will absorb radiation at this resonance frequency ν* - and this frequency, as you might have already guessed, corresponds to ΔE, the energy difference between the proton's two spin states. With the strong magnetic fields generated by the superconducting magnets used in modern NMR instruments, the resonance frequency for protons falls within the radio-wave range, anywhere from 100 MHz to 800 MHz depending on the strength of the magnet.

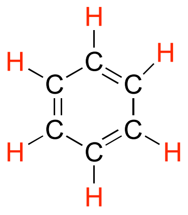
Section 5.2: Chemical equivalence

If all protons in all organic molecules had the same resonance frequency in an external magnetic field of a given strength, the information in the previous paragraph would be interesting from a theoretical standpoint, but would not be terribly useful to organic chemists. Fortunately for us, however, resonance frequencies are not uniform for all protons in a molecule. *In an external magnetic field of a given strength, protons in different locations in a molecule have different resonance frequencies, because they are in non-identical electronic environments.* In methyl acetate, for example, there are two 'sets' of protons. The three protons labeled H_a have a different - and easily distinguishable – resonance frequency than the three H_b protons, because the two sets of protons are in non-identical environments: they are, in other words, chemically nonequivalent.

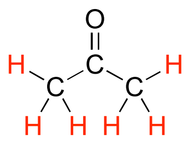


On the other hand, the three H_a protons are all in the same electronic environment, and are chemically equivalent to one another. They have identical resonance frequencies. The same can be said for the three H_b protons.

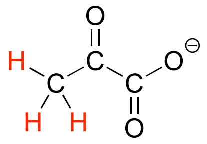
The ability to recognize chemical equivalancy and nonequivalency among atoms in a molecule will be central to understanding NMR. In each of the molecules below, all protons are chemically equivalent, and therefore will have the same resonance frequency in an NMR experiment.



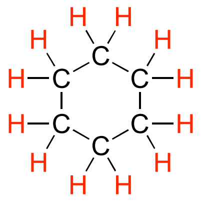
benzene



acetone

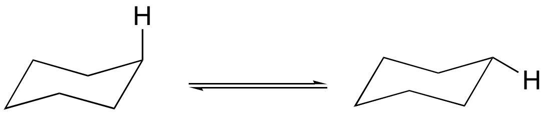


pyruvate



cyclohexane

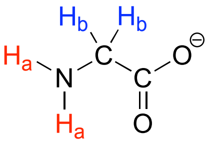
You might expect that the equitorial and axial hydrogens in cyclohexane would be non-equivalent, and would have different resonance frequencies. In fact, an axial hydrogen *is* in a different electronic environment than an equitorial hydrogen. Remember, though, that the molecule rotates rapidly between its two chair conformations, meaning that any given hydrogen is rapidly moving back and forth between equitorial and axial positions. It turns out that, except at extremely low temperatures, this rotational motion occurs on a time scale that is much faster than the time scale of an NMR experiment.



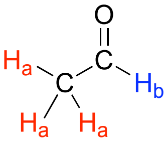
ring-flip process is *fast* compared to the NMR time-scale

In this sense, NMR is like a camera that takes photographs of a rapidly moving object with a slow shutter speed - the result is a blurred image. In NMR terms, this means that all 12 protons in cyclohexane are equivalent.

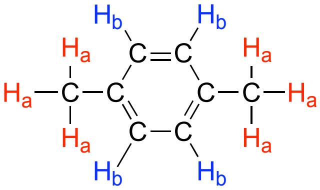
Each the molecules in the next figure contains *two* sets of protons, just like our previous example of methyl acetate, and again in each case the resonance frequency of the H_a protons will be different from that of the H_b protons.



glycine



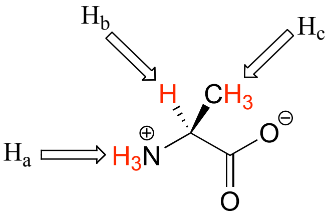
acetaldehyde



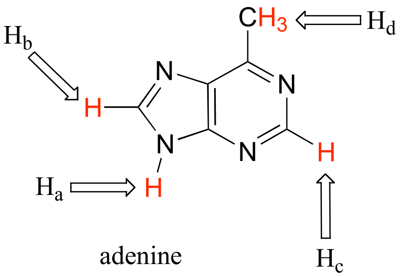
para-xylene

Notice how the symmetry of *para*-xylene results in there being only two different sets of protons.

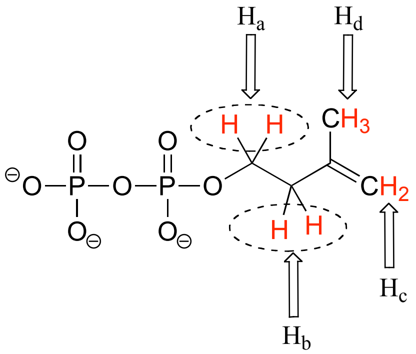
Most organic molecules have several sets of protons in different chemical environments, and each set, in theory, will have a different resonance frequency in ^{1}H-NMR spectroscopy.



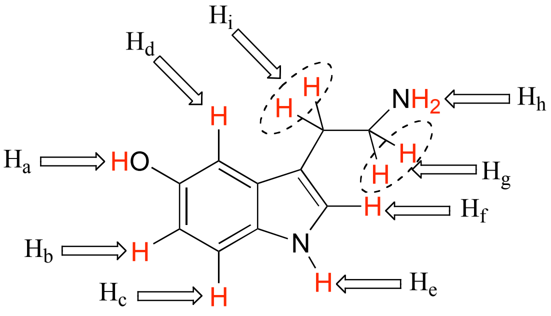
alanine
(an amino acid)



adenine
(a DNA/RNA base)

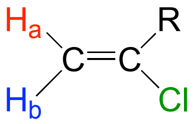


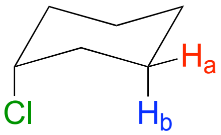
isopentenyl diphosphate
(a lipid building block)



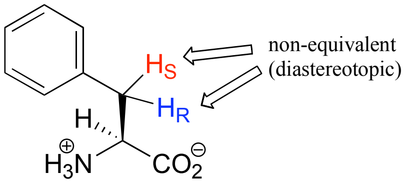
seratonin (a neurotransmitter)

When stereochemistry is taken into account, the issue of equivalence vs nonequivalence in NMR starts to get a little more complicated. It should be fairly intuitive that hydrogens on different sides of asymmetric ring structures and double bonds are in different electronic environments, and thus are non-equivalent and have different resonance frequencies. In the alkene and cyclohexene structures below, for example, H_a is *trans* to the chlorine substituent, while Hb is *cis* to chlorine.



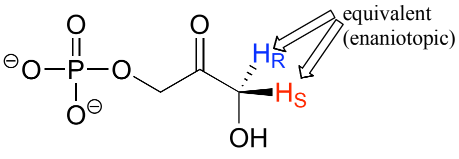


What is not so intuitive is that diastereotopic hydrogens (section 3.10) on chiral molecules are also non-equivalent:

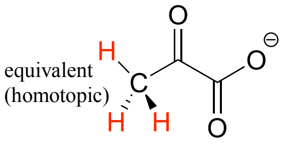


phenylalanine

However, enantiotopic and homotopic hydrogens are chemically equivalent.



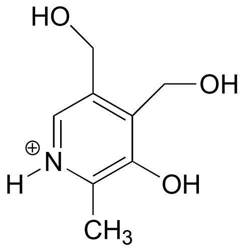
dihydroxyacetone
phosphate



pyruvate

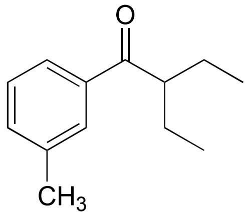
Exercise 5.1: How many different sets of protons do the following molecules contain? (count diastereotopic protons as non-equivalent).

a)



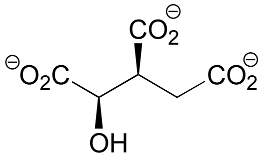
pyridoxine
(vitamin B_6)

b)



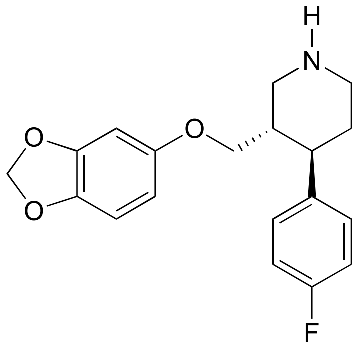
DEET
(an insect repellent)

c)



isocitrate
(a citric acid cycle intermediate)

d)

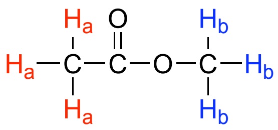


paroxetine
(an antianxietry drug, trade name Paxil)

Section 5.3: The NMR experiment

5.3A: The basics of an NMR experiment

Given that chemically nonequivalent protons have different resonance frequencies in the same applied magnetic field, we can see how NMR spectroscopy can provide us with useful information about the structure of an organic molecule. A full explanation of how a modern NMR instrument functions is beyond the scope of this text, but in very simple terms, here is what happens. First, a sample compound (we'll use methyl acetate) is placed inside a very strong applied magnetic field (B_0).

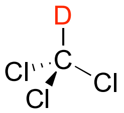


$$\mathrm{H_a{-}\overset{H_a}{\underset{H_a}{C}}{-}\overset{O}{\overset{\|}{C}}{-}O{-}\overset{H_b}{\underset{H_b}{C}}{-}H_b}$$

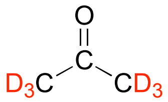
methyl acetate

All of the protons begin to precess: the H_a protons at precessional frequency ω_a, the H_b protons at ω_b. At first, the magnetic moments of (slightly more than) half of the protons are aligned with B_0, and half are aligned against B_0. Then, the sample is hit with electromagnetic radiation in the radio frequency range. The two specific frequencies which match ω_a and ω_b (i.e. the resonance frequencies) cause those H_a and H_b protons which are aligned *with* B_0 to 'flip' so that they are now aligned *against* B_0. In doing so, the protons absorb radiation at the two resonance frequencies. The NMR instrument records which frequencies were absorbed, as well as the intensity of each absorbance.

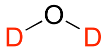
In most cases, a sample being analyzed by NMR is in solution. If we use a common laboratory solvent (diethyl ether, acetone, dichloromethane, ethanol, water, etc.) to dissolve our NMR sample, however, we run into a problem – there many more solvent protons in solution than there are sample protons, so the signals from the sample protons will be overwhelmed. To get around this problem, we use special NMR solvents in which all protons have been replaced by deuterium. Recall that deuterium is NMR-active, but its resonance frequency is very different from that of protons, and thus it is `invisible` in ^{1}H-NMR. Some common NMR solvents are shown below.



deuterated chloroform
($CDCl_3$)



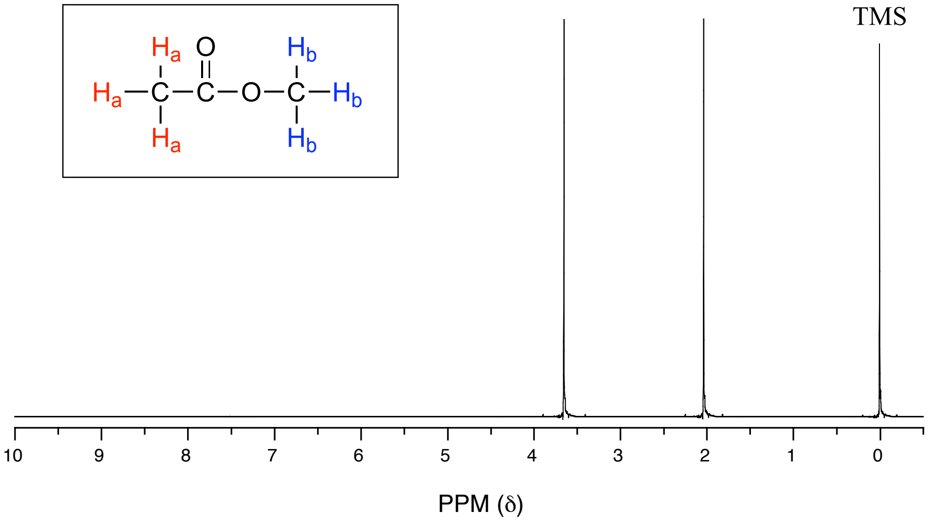
d_6-acetone



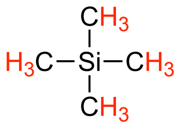
D_2O

5.3B: The chemical shift

Let's look at an actual ^{1}H-NMR plot for methyl acetate. Just as in IR and UV-vis spectroscopy, the vertical axis corresponds to intensity of absorbance, the horizontal axis to frequency (typically the vertical axis is not shown in an NMR spectrum).



We see three absorbance signals: two of these correspond to H_a and H_b, while the peak at the far right of the spectrum corresponds to the 12 chemically equivalent protons in tetramethylsilane (TMS), a standard reference compound that was added to our sample.



$$H_3C-\overset{\overset{CH_3}{|}}{\underset{\underset{CH_3}{|}}{Si}}-CH_3$$

tetramethylsilane (TMS)

You may be wondering about a few things at this point - why is TMS necessary, and what is the meaning of the `ppm` label on the horizontal axis? Shouldn't the frequency units be in Hz? Keep in mind that NMR instruments of many different applied field strengths are used in organic chemistry laboratories, and that the proton's resonance frequency range depends on the strength of the applied field. The spectrum above was generated on an instrument with an applied field of approximately 7.1 Tesla, at which strength protons resonate in the neighborhood of 300 million Hz (chemists refer to this as a 300 MHz instrument). If our colleague in another lab takes the NMR spectrum of the same molecule using an instrument with a 2.4 Tesla magnet, the protons will resonate at around 100 million Hz (so we'd call this a 100 MHz instrument). It would be inconvenient and confusing to always have to convert NMR data according to the field strength of the instrument used. Therefore, chemists report resonance frequencies not as absolute values in Hz, but rather as values *relative to a common standard*, generally the signal generated by the protons in TMS. This is where the ppm – parts per million – term comes in. Regardless of the magnetic field strength of the instrument being used, the resonance frequency of the 12 equivalent protons in TMS is defined as a zero point. The resonance frequencies of protons in the sample molecule are then reported in terms of how much higher they are, in ppm, relative to the TMS signal (almost all protons in organic molecules have a higher resonance frequency than those in TMS, for reasons we shall explore quite soon).

The two proton groups in our methyl acetate sample are recorded as resonating at frequencies 2.05 and 3.67 ppm higher than TMS. One-millionth (1.0 ppm) of 300 MHz is 300 Hz. Thus 2.05 ppm, on this instrument, corresponds to 615 Hz, and 3.67 ppm corresponds to 1101 Hz. If the TMS protons observed by our 7.1 Tesla instrument resonate at exactly 300,000,000 Hz, this means that the protons in our ethyl acetate samples are resonating at 300,000,615 and 300,001,101 Hz, respectively. Likewise, if the TMS protons in our colleague's 2.4 Tesla instrument resonate at exactly 100 MHz, the methyl acetate protons in her sample resonate at 100,000,205 and 100,000,367 Hz (on the 100 MHz instrument, 1.0 ppm corresponds to 100 Hz). The absolute frequency values in each case are not very useful – they will vary according to the instrument used – but the *difference* in resonance frequency from the TMS standard, expressed in parts per million, should be the same regardless of the instrument.

Expressed this way, the resonance frequency for a given proton in a molecule is called its **chemical shift**. A frequently used symbolic designation for chemical shift in ppm is the lower-case greek letter δ (*delta*). Most protons in organic compounds have chemical shift values between 0 and 12 ppm from TMS, although values below zero and above 12 are occasionally observed. By convention, the left-hand side of an NMR spectrum (higher chemical shift) is called **downfield**, and the right-hand direction is called **upfield**.

In our methyl acetate example we included for illustrative purposes a small amount of TMS standard directly in the sample, as was the common procedure for determining the zero point with older NMR instruments. That practice is generally no longer necessary, as modern NMR instruments are designed to use the deuterium signal from the solvent as a standard reference point, then to extrapolate the 0 ppm baseline that corresponds to the TMS proton signal (in an applied field of 7.1 Tesla, the deuterium atom in $CDCl_3$ resonates at 32 MHz, compared to 300 MHz for the protons in TMS). In the remaining NMR spectra that we will see in this text we will not see an actual TMS signal, but we can always assume that the 0 ppm point corresponds to where the TMS protons *would* resonate if they were present.

Exercise 5.2: A proton has a chemical shift (relative to TMS) of 4.56 ppm.

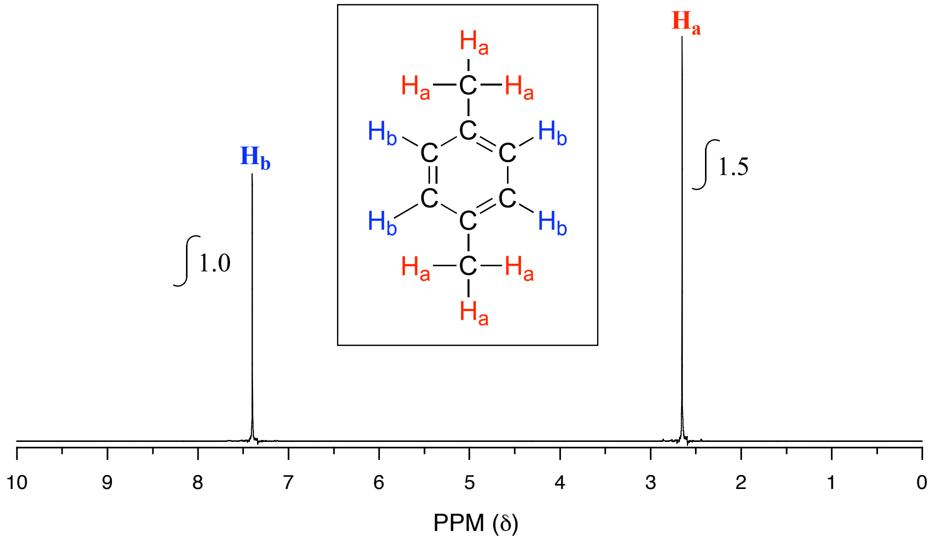
a) What is its chemical shift, expressed in Hz, in a 300 MHz instrument? On a 200 MHz instrument?
b) What is its resonance frequency, expressed in Hz, in a 300 MHz instrument? On a 200 MHz instrument?

(Assume that in these instruments, the TMS protons resonate at exactly 300 or 200 MHz, respectively)

5.3C: Signal integration

The computer in an NMR instrument can be instructed to automatically integrate the area under a signal or group of signals. This is very useful, because *in ^{1}H-NMR spectroscopy the area under a signal is proportional to the number of hydrogens to which the peak corresponds*. The two signals in the methyl acetate spectrum, for example, integrate to approximately the same area, because they both correspond to a set of three equivalent protons.

Take a look next at the spectrum of *para*-xylene (IUPAC name 1,4-dimethylbenzene):



This molecule has two sets of protons: the six methyl (H_a) protons and the four aromatic (H_b) protons. When we instruct the instrument to integrate the areas under the two signals, we find that the area under the peak at 2.6 ppm is 1.5 times greater than the area under the peak at 7.4 ppm. This (along with the actual chemical shift values, which we'll discuss soon) tells us which set of protons corresponds to which NMR signal.

The integration function can also be used to determine the relative amounts of two or more compounds in a *mixed* sample. If we have a sample that is a 50:50 (mole/mole) mixture of benzene and acetone, for example, the acetone signal should integrate to the same value as the benzene sample, because both signals represent six equivalent protons. If we have a 50:50 mixture of acetone and cyclopentane, on the other hand, the ratio of the acetone peak area to the cylopentane peak area will be 3:5 (or 6:10), because the cyclopentane signal represents ten protons.

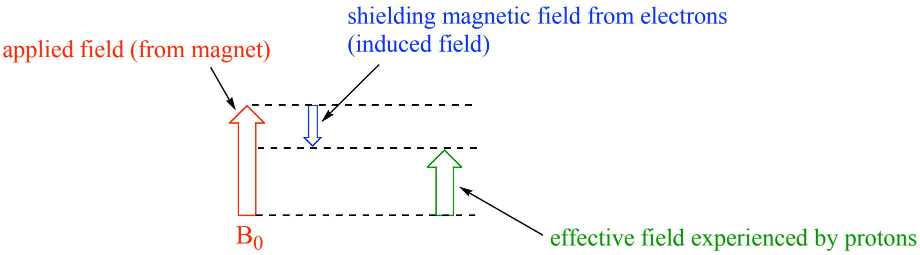
Exercise 5.3: You take a ^{1}H-NMR spectrum of a mixed sample of acetone ($CH_3(CO)CH_3$) and dichloromethane (CH_2Cl_2). The integral ratio of the two signals (acetone : dichloromethane) is 2.3 to 1. What is the molar ratio of the two compounds in the sample?

Exercise 5.4: You take the ^{1}H-NMR spectrum of a mixed sample of 36% *para*-xylene and 64% acetone in $CDCl_3$ solvent (structures are shown earlier in this chapter). How many peaks do you expect to see? What is the expected ratio of integration values for these peaks? (set the acetone peak integration equal to 1.0)

Section 5.4: The basis for differences in chemical shift

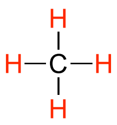
5.4A: Diamagnetic shielding and deshielding

We come now to the question of *why* nonequivalent protons have different chemical shifts. The chemical shift of a given proton is determined primarily by its immediate electronic environment. Consider the methane molecule (CH_4), in which the protons have a chemical shift of 0.23 ppm. The valence electrons around the methyl carbon, when subjected to B_0, are induced to circulate and thus generate their own very small magnetic field that *opposes* B_0. This **induced field**, to a small but significant degree, *shields* the nearby protons from experiencing the full force of B_0, an effect known as **local diamagnetic shielding**. The methane protons therefore do not experience the full force of B_0 - what they experience is called B_{eff}, or the **effective field**, which is slightly *weaker* than B_0.

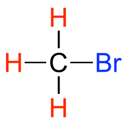

shielding magnetic field from electrons
(induced field)
applied field (from magnet)
B_0
effective field experienced by protons

Therefore, their resonance frequency is slightly lower than what it would be if they did not have electrons nearby to shield them.

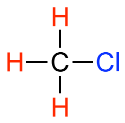
Now consider methyl fluoride, CH_3F, in which the protons have a chemical shift of 4.26 ppm, significantly higher than that of methane. This is caused by something called the **deshielding effect**. Because fluorine is more electronegative than carbon, it pulls valence electrons away from the carbon, effectively *decreasing* the electron density around each of the protons. For the protons, lower electron density means less diamagnetic shielding, which in turn means a greater overall exposure to B_0, a stronger B_{eff}, and a higher resonance frequency. Put another way, the fluorine, by pulling electron density away from the protons, is *deshielding* them, leaving them more exposed to B_0. As the electronegativity of the substituent increases, so does the extent of deshielding, and so does the chemical shift. This is evident when we look at the chemical shifts of methane and three halomethane compounds (remember that electronegativity increases as we move up a column in the periodic table).



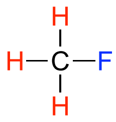
0.23 ppm



2.68 ppm

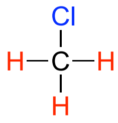


3.05 ppm

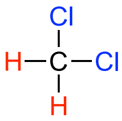


4.26 ppm

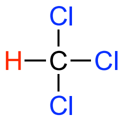
To a large extent, then, we can predict trends in chemical shift by considering how much deshielding is taking place near a proton. The chemical shift of trichloromethane is, as expected, higher than that of dichloromethane, which is in turn higher than that of chloromethane.



3.05 ppm

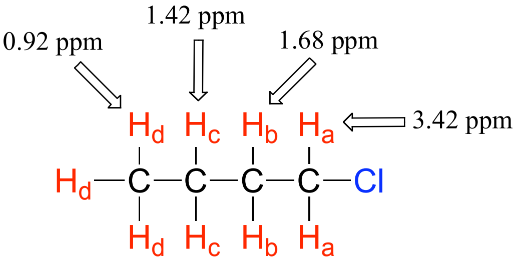


5.30 ppm



7.27 ppm

The deshielding effect of an electronegative substituent diminishes sharply with increasing distance:



The presence of an electronegative oxygen, nitrogen, sulfur, or sp^2-hybridized carbon also tends to shift the NMR signals of nearby protons slightly downfield:

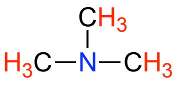
H_3C-CH_3

$H_3C-O-CH_3$

δ 3.42 ppm

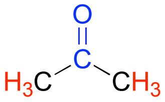
$H_3C-S-CH_3$

δ 2.12 ppm

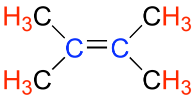


$(CH_3)N(CH_3)_2$

δ 2.12 ppm



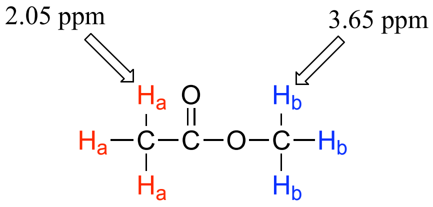
2.16 ppm



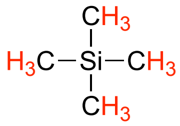
1.64 ppm

Table 2 in the tables section at the end of the text lists typical chemical shift values for protons in different chemical environments.

Armed with this information, we can finally assign the two peaks in the the ^{1}H-NMR spectrum of methyl acetate that we saw a few pages back. The signal at 3.65 ppm corresponds to the methyl ester protons (H_b), which are deshielded by the adjacent oxygen atom. The upfield signal at 2.05 ppm corresponds to the acetate protons (H_a), which is deshielded - but to a lesser extent - by the adjacent carbonyl group.

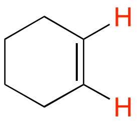


Finally, a note on the use of TMS as a standard in NMR spectroscopy: one of the main reasons why the TMS proton signal was chosen as a zero-point is that the TMS protons are highly shielded: silicon is slightly *less* electronegative than carbon, and therefore *donates* some additional shielding electron density. Very few organic molecules contain protons with chemical shifts that are negative relative to TMS.

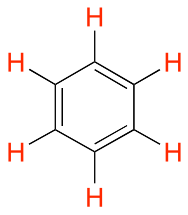


5.4B: Diamagnetic anisotropy

Some protons resonate much further downfield than can be accounted for simply by the deshielding effect of nearby electronegative atoms. Vinylic protons (those directly bonded to an alkene carbon) and aromatic (benzylic) protons are dramatic examples.



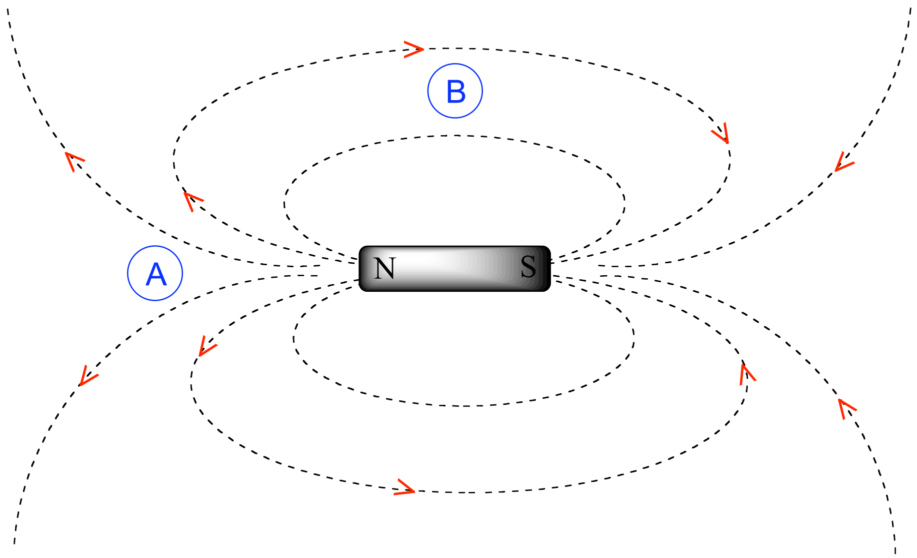
5.66 ppm



7.34 ppm

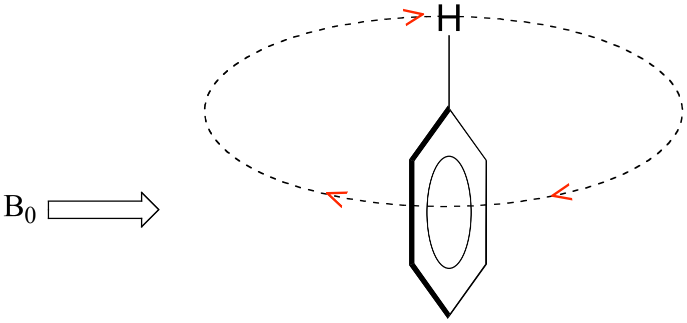
We'll consider the aromatic proton first. Recall that in benzene and many other aromatic structures, a sextet of π electrons is delocalized around the ring. When the molecule is exposed to B_0, these π electrons begin to circulate in a **ring current**, generating their own induced magnetic field that opposes B_0. In this case, however, the induced field of the π electrons does not shield the benzylic protons from B_0 as you might expect– rather, it causes the protons to experience a *stronger* magnetic field in the direction of B_0 – in other words, it *adds* to B_0 rather than subtracting from it.

To understand how this happens, we need to understand the concept of **diamagnetic anisotropy** (anisotropy means `non-uniformity`). So far, we have been picturing magnetic fields as being oriented in a uniform direction. This is only true over a small area. If we step back and take a wider view, however, we see that the lines of force in a magnetic field are actually anisotropic. They start in the 'north' direction, then loop around like a snake biting its own tail.



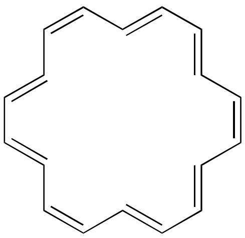
If we are at point A in the figure above, we feel a magnetic field pointing in a northerly direction. If we are at point B, however, we feel a field pointing to the south.

In the induced field generated by the aromatic ring current, the benzylic protons are at the equivalent of ‘point B’ – this means that the induced current in this region of space is oriented in the *same* direction as B_0.



In total, the benzylic protons are subjected to three magnetic fields: the applied field (B_0) and the induced field from the π electrons pointing in one direction, and the induced field of the non-aromatic electrons pointing in the opposite (shielding) direction. The end result is that benzylic protons, due to the anisotropy of the induced field generated by the π ring current, appear to be highly deshielded. Their chemical shift is far downfield, in the 6.5-8 ppm region.

Exercise 5.5: The ^{1}H-NMR spectrum of [18] annulene has two peaks, at 8.9 ppm and -1.8 ppm (upfield of TMS!) with an integration ratio of 2:1. Explain the unusual chemical shift of the latter peak.



[18] annulene

Diamagnetic anisotropy is also responsible for the downfield chemical shifts of vinylic protons and aldehyde protons (4.5-6.5 ppm and 9-10 ppm, respectively). These groups are not aromatic and thus do not generate ring currents as does benzene, but the π electrons circulate in such a way as to generate a magnetic field that *adds* to B_0 in the regions of space occupied by the protons.

5.4C: Hydrogen-bonded protons

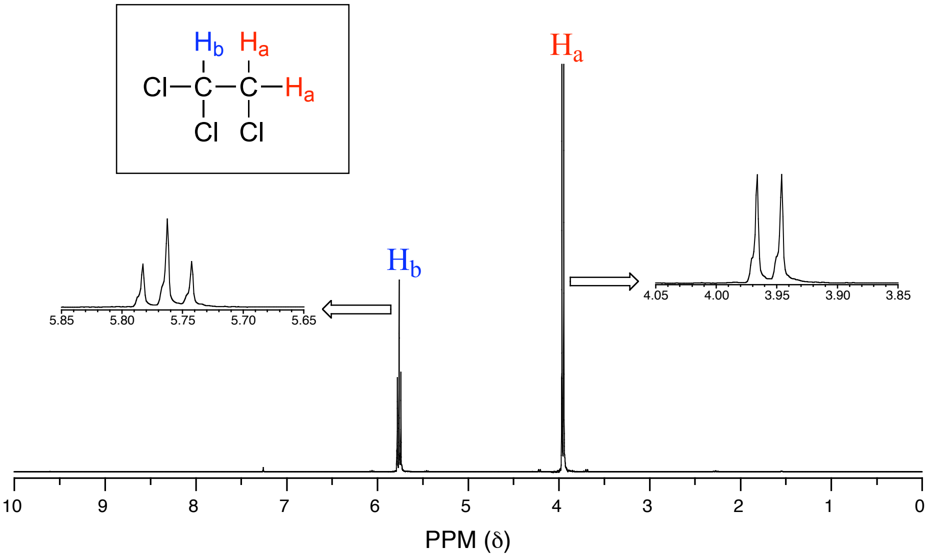
Protons that are directly bonded to oxygen, nitrogen, and sulfur have chemical shifts that can vary widely depending on solvent and concentration. This is because these protons can participate to varying degrees in hydrogen-bonding interactions, and hydrogen bonding greatly influences the electron density around the proton. Signals for hydrogen-bonding protons also tend to be broader than those of hydrogens bonded to carbon, a phenomenon that is also due to hydrogen bonding.

Section 5.5: Spin-spin coupling

5.5A: The source of spin-spin coupling

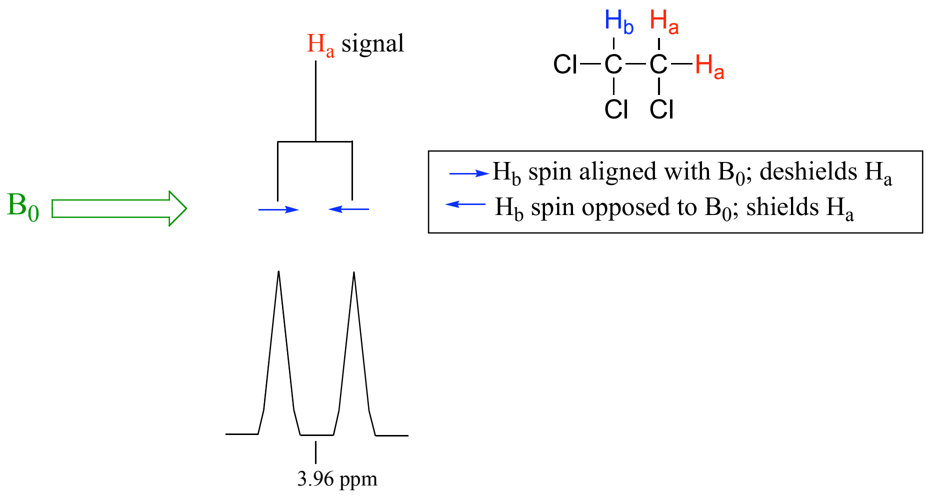
The ^{1}H-NMR spectra that we have seen so far (of methyl acetate and *para*-xylene) are somewhat unusual in the sense that in both of these molecules, each set of protons generates a single NMR signal. In fact, the ^{1}H-NMR spectra of most organic molecules contain proton signals that are 'split' into two or more sub-peaks. Rather than being a complication, however, this splitting behavior actually provides us with more information about our sample molecule.

Consider the spectrum for 1,1,2-trichloroethane. In this and in many spectra to follow, we show enlargements of individual signals so that the signal splitting patterns are recognizable.

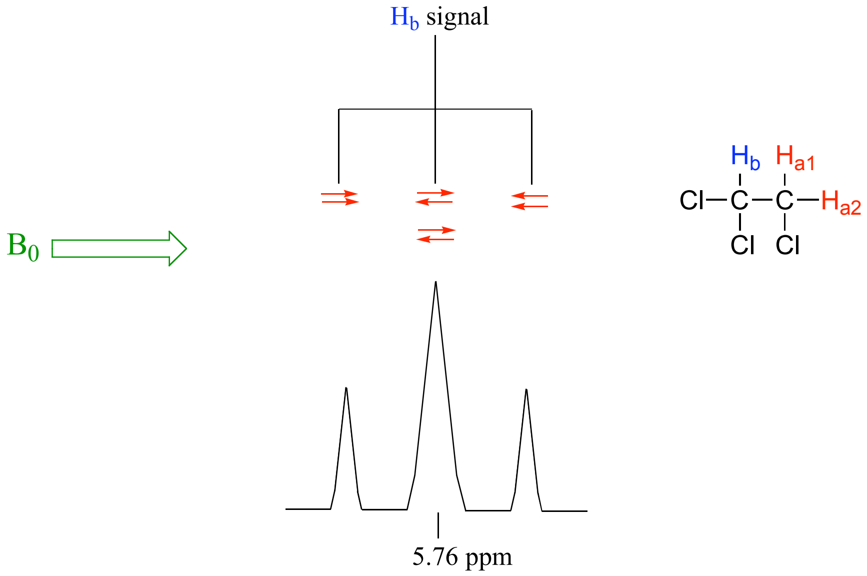


The signal at 3.96 ppm, corresponding to the two H_a protons, is split into two subpeaks of equal height (and area) – this is referred to as a **doublet**. The H_b signal at 5.76 ppm, on the other hand, is split into three sub-peaks, with the middle peak higher than the two outside peaks - if we were to integrate each subpeak, we would see that the area under the middle peak is twice that of each of the outside peaks. This is called a **triplet**.

The source of signal splitting is a phenomenon called **spin-spin coupling**, a term that describes the magnetic interactions between neighboring, non-equivalent NMR-active nuclei. In our 1,1,2 trichloromethane example, the H_a and H_b protons are spin-coupled to each other. Here's how it works, looking first at the H_a signal: in addition to being shielded by nearby valence electrons, each of the H_a protons is also influenced by the small magnetic field generated by H_b next door (remember, each spinning proton is like a tiny magnet). The magnetic moment of H_b will be aligned *with* B_0 in (slightly more than) half of the molecules in the sample, while in the remaining half of the molecules it will be opposed to B_0. The B_{eff} 'felt' by H_a is a slightly weaker if H_b is aligned against B_0, or slightly stronger if H_b is aligned with B_0. In other words, in half of the molecules H_a is *shielded* by H_b (thus the NMR signal is shifted slightly upfield) and in the other half H_a is *deshielded* by H_b (and the NMR signal shifted slightly downfield). What would otherwise be a single H_a peak has been split into two sub-peaks (a doublet), one upfield and one downfield of the original signal. These ideas an be illustrated by a **splitting diagram**, as shown below.

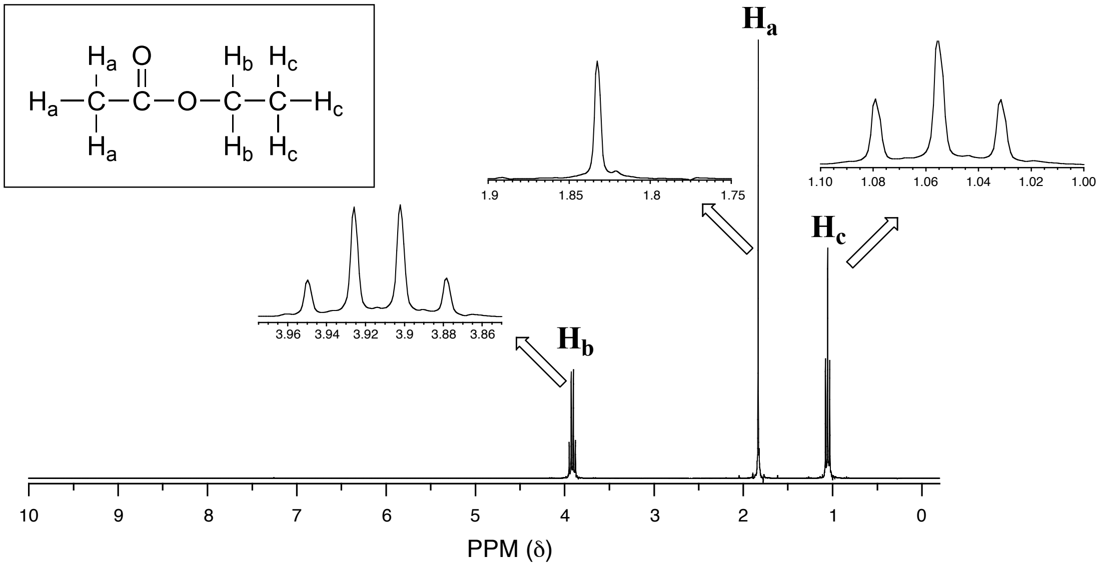

H_a signal
H_b H_a
Cl—C—C—H_a
Cl Cl
B_0
H_b spin aligned with B_0; deshields H_a
H_b spin opposed to B_0; shields H_a
3.96 ppm

Now, let's think about the H_b signal. The magnetic environment experienced by H_b is influenced by the fields of both neighboring H_a protons, which we will call H_{a1} and H_{a2}. There are four possibilities here, each of which is equally probable. First, the magnetic fields of both H_{a1} and H_{a2} could be aligned with B_0, which would deshield H_b, shifting its NMR signal slightly downfield. Second, both the H_{a1} and H_{a2} magnetic fields could be aligned opposed to B_0, which would shield H_b, shifting its resonance signal slightly upfield. Third and fourth, H_{a1} could be with B_0 and H_{a2} opposed, or H_{a1} opposed to B_0 and H_{a2} with B_0. In each of the last two cases, the shielding effect of one H_a proton would cancel the deshielding effect of the other, and the chemical shift of H_b would be unchanged.



So in the end, the signal for H_b is a **triplet**, with the middle peak twice as large as the two outer peaks because there are *two* ways that H_{a1} and H_{a2} can cancel each other out.

Now, consider the spectrum for ethyl acetate:



We see an unsplit **'singlet'** peak at 1.833 ppm that corresponds to the acetyl (H_a) hydrogens – this is similar to the signal for the acetate hydrogens in methyl acetate that we considered earlier. This signal is unsplit because there are no adjacent hydrogens on the molecule. The signal at 1.055 ppm for the H_c hydrogens is split into a triplet by the two H_b hydrogens next door. The explanation here is the same as the explanation for the triplet peak we saw previously for 1,1,2-trichloroethane.

The H_b hydrogens give rise to a **quartet** signal at 3.915 ppm – notice that the two middle peaks are taller then the two outside peaks. This splitting pattern results from the spin-coupling effect of the *three* H_c hydrogens next door, and can be explained by an analysis similar to that which we used to explain the doublet and triplet patterns.

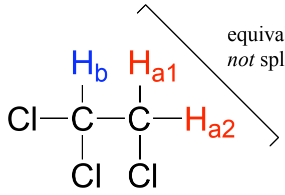
Exercise 5.6:

a) Explain, using left and right arrows to illustrate the possible combinations of nuclear spin states for the H_c hydrogens, why the H_b signal in ethyl acetate is split into a quartet.

b) The integration ratio of doublets is 1:1, and of triplets is 1:2:1. What is the integration ratio of the H_b quartet in ethyl acetate? (Hint – use the illustration that you drew in part a to answer this question.)

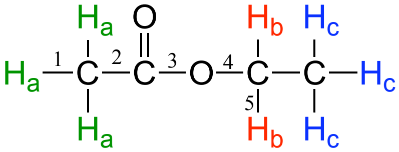
By now, you probably have recognized the pattern which is usually referred to as the ***n* + 1 rule**: if a set of hydrogens has *n* neighboring, non-equivalent hydrogens, it will be split into *n* + 1 subpeaks. Thus the two H_b hydrogens in ethyl acetate split the H_c signal into a triplet, and the three H_c hydrogens split the H_b signal into a quartet. This is very useful information if we are trying to determine the structure of an unknown molecule: if we see a triplet signal, we know that the corresponding hydrogen or set of hydrogens has two `neighbors`. When we begin to determine structures of unknown compounds using ^{1}H-NMR spectral data, it will become more apparent how this kind of information can be used.

Three important points need to be emphasized here. First, signal splitting only occurs between non-equivalent hydrogens – in other words, H_{a1} in 1,1,2-trichloroethane is *not* split by H_{a2}, and vice-versa.



equivalent hydrogens do *not* split each other's signals

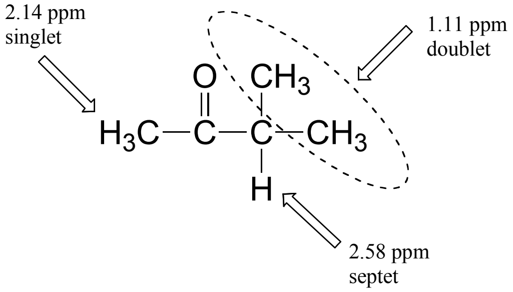
Second, splitting occurs primarily between hydrogens that are separated by three bonds. This is why the H_a hydrogens in ethyl acetate form a singlet– the nearest hydrogen neighbors are five bonds away, too far for coupling to occur.

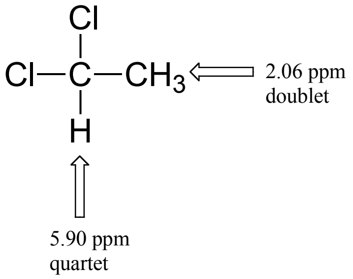


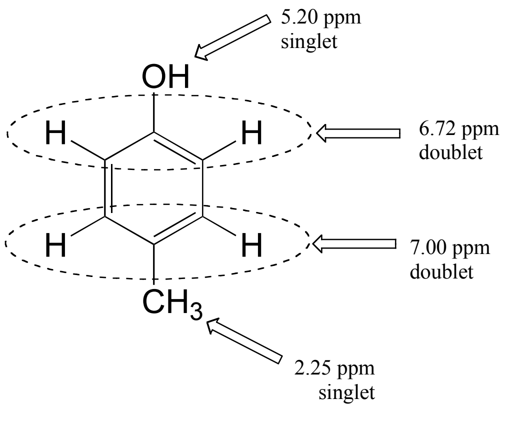
H_a protons are 5 bonds away from H_b - no coupling

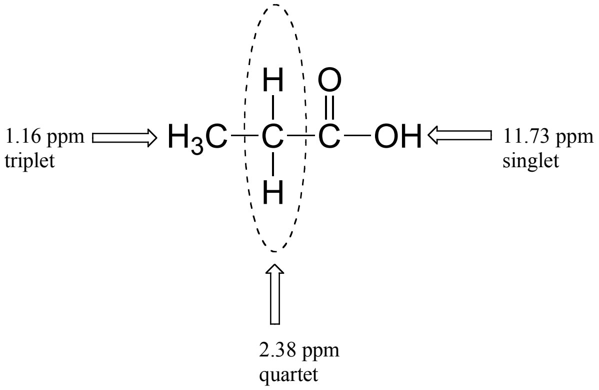
Occasionally we will see four-bond and even 5-bond splitting, but in these cases the magnetic influence of one set of hydrogens on the other set is much more subtle than what we typically see in three-bond splitting (more details about how we quantify coupling interactions is provided in section 5.5B). Finally, splitting is most noticeable with hydrogens bonded to carbon. Hydrogens that are bonded to heteroatoms (alcohol or amino hydrogens, for example) are coupled weakly - or not at all - to their neighbors. This has to do with the fact that these protons exchange rapidly with solvent or other sample molecules.

Below are a few more examples of chemical shift and splitting pattern information for some relatively simple organic molecules.

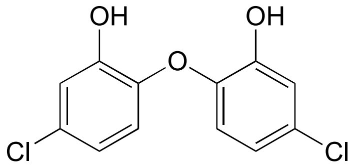




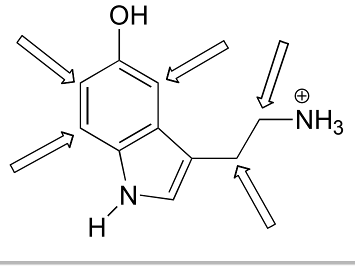




Exercise 5.7: How many proton signals would you expect to see in the ^{1}H-NMR spectrum of triclosan (a common antimicrobial agent found in detergents)? For each of the proton signals, predict the splitting pattern. Assume that you see only 3-bond coupling.

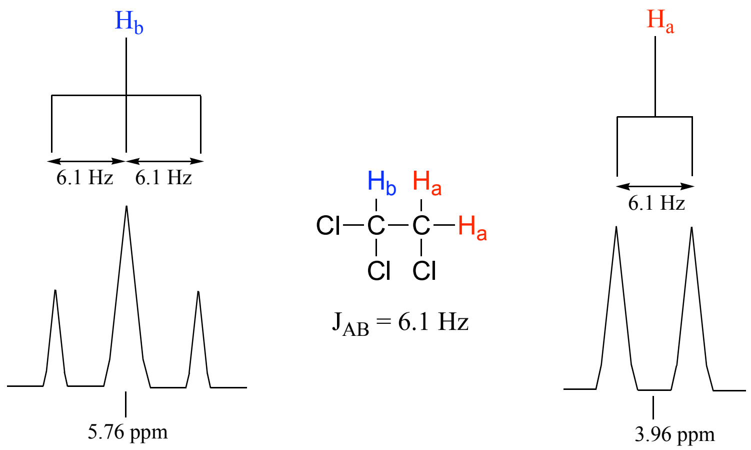


Exercise 5.8: Predict the splitting pattern for the ^{1}H-NMR signals corresponding to the protons at the locations indicated by arrows (the structure is that of the neurotransmitter serotonin).



5.5B: Coupling constants

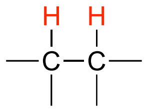
Chemists quantify the spin-spin coupling effect using something called the **coupling constant**, which is abbreviated with the capital letter J. The coupling constant is simply the difference, expressed in Hz, between two adjacent sub-peaks in a split signal. For our doublet in the 1,1,2-trichloroethane spectrum, for example, the two subpeaks are separated by 6.1 Hz, and thus we write $^3J_{a\text{-}b} = 6.1$ Hz.



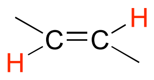
The superscript 3 tells us that this is a three-bond coupling interaction, and the a-b subscript tells us that we are talking about coupling between H_a and H_b. Unlike the chemical shift value, *the coupling constant, expressed in Hz, is the same regardless of the applied field strength of the NMR magnet*. This is because the strength of the magnetic moment of a neighboring proton, which is the source of the spin-spin coupling phenomenon, does *not* depend on the applied field strength.

When we look closely at the triplet signal in 1,1,2-trichloroethane, we see that the coupling constant - the `gap` between subpeaks - is 6.1 Hz, the same as for the doublet. This is an important concept! The coupling constant $^3J_{a\text{-}b}$ quantifies the magnetic interaction between the H_a and H_b hydrogen sets, and *this interaction is of the same magnitude in either direction*. In other words, H_a influences H_b to the same extent that H_b influences H_a. When looking at more complex NMR spectra, this idea of **reciprocal coupling constants** can be very helpful in identifying the coupling relationships between proton sets.

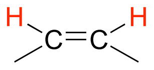
Coupling constants between proton sets on neighboring sp^3-hybridized carbons is typically in the region of 6-8 Hz. With protons bound to sp^2-hybridized carbons, coupling constants can range from 0 Hz (no coupling at all) to 18 Hz, depending on the bonding arrangement.



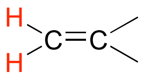
^3J = 6-8 Hz



vinylic *trans*:
^3J = 11-18 Hz

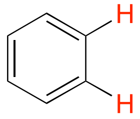


vinylic *cis*:
^3J = 6-15 Hz

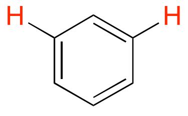


vinylic *gem*:
^2J = 0-5 Hz

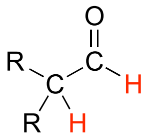
For vinylic hydrogens in a *trans* configuration, we see coupling constants in the range of ^3J = 11-18 Hz, while *cis* hydrogens couple in the ^3J = 6-15 Hz range. The 2-bond coupling between hydrogens bound to the same alkene carbon (referred to as geminal hydrogens) is very fine, generally 5 Hz or lower. *Ortho* hydrogens on a benzene ring couple at 6-10 Hz, while 4-bond coupling of up to 4 Hz is sometimes seen between *meta* hydrogens.



ortho-benzylic:
^{3}J = 6-10 Hz



meta-benzylic:
^{4}J = 0-4 Hz



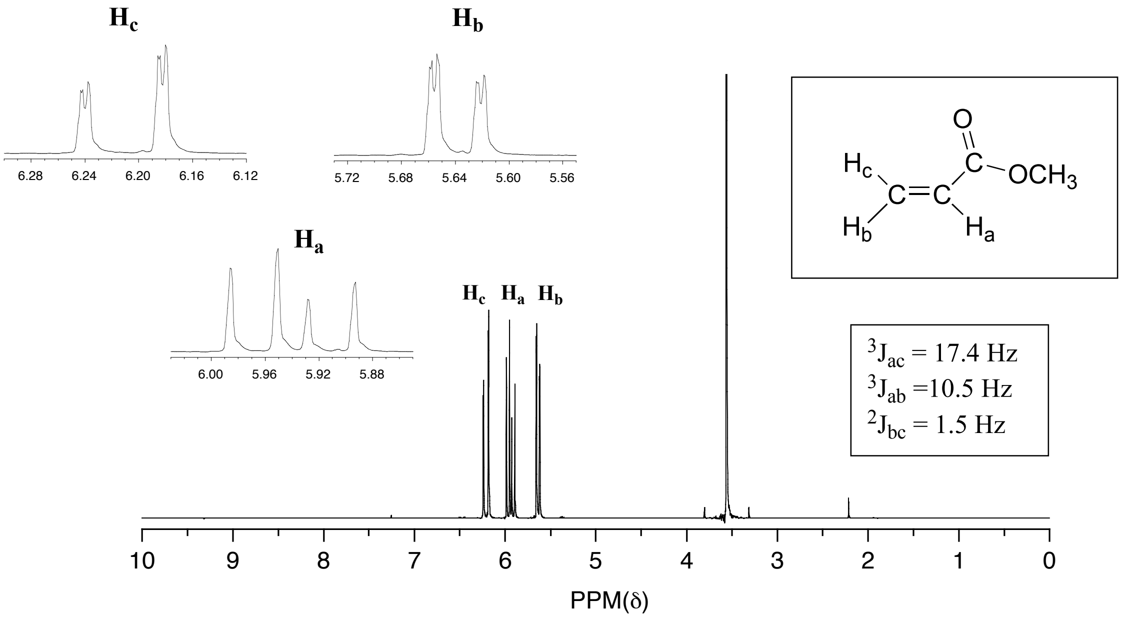
aldehyde:
^{3}J = 2-3 Hz

Fine (2-3 Hz) coupling is often seen between an aldehyde proton and a three-bond neighbor.

Table 4, in the table section at the end of the text, summarizes the most important coupling constant values.

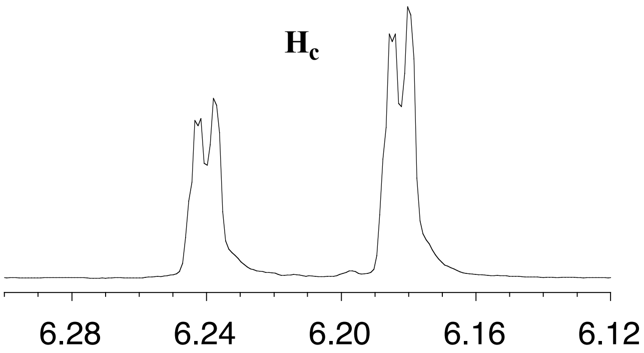
5.5C: Complex coupling

In all of the examples of spin-spin coupling that we have seen so far, the observed splitting has resulted from the coupling of one set of hydrogens to *just one* neighboring set of hydrogens. When a set of hydrogens is coupled to *two or more* sets of nonequivalent neighbors, the result is a phenomenon called **complex coupling**. A good illustration is provided by the ^{1}H-NMR spectrum of methyl acrylate:

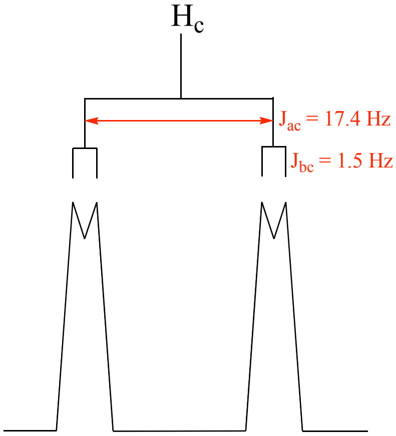


$^{3}J_{ac}$ = 17.4 Hz
$^{3}J_{ab}$ =10.5 Hz
$^{2}J_{bc}$ = 1.5 Hz

First, let's first consider the H$_c$ signal, which is centered at 6.21 ppm. Here is a closer look:

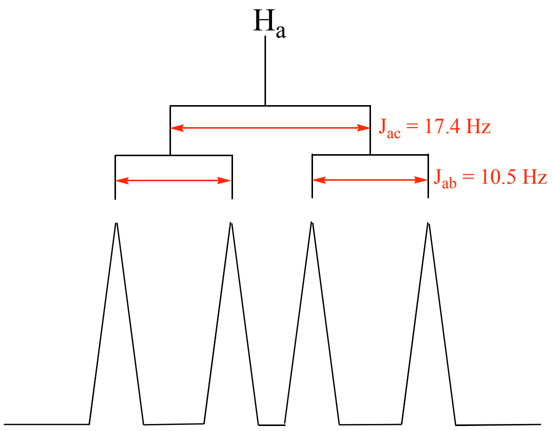


With this enlargement, it becomes evident that the Hc signal is actually composed of four sub-peaks. Why is this? H_c is coupled to both H_a and H_b , but with *two different coupling constants*. Once again, a splitting diagram can help us to understand what we are seeing. H_a is *trans* to H_c across the double bond, and splits the H_c signal into a doublet with a coupling constant of $^3J_{ac}$ = 17.4 Hz. In addition, each of these H_c doublet sub-peaks is split again by H_b (*geminal* coupling) into two more doublets, each with a much smaller coupling constant of $^2J_{bc}$ = 1.5 Hz.

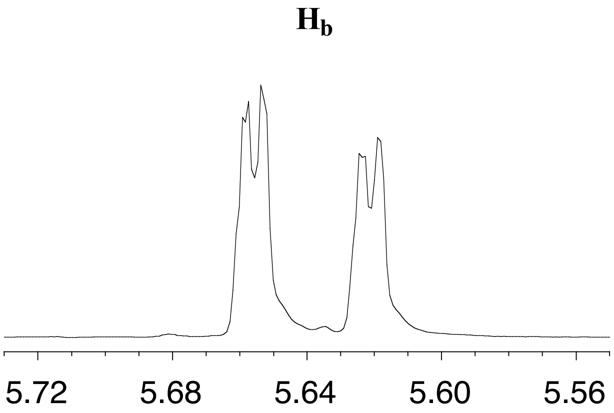


The result of this `double splitting` is a pattern referred to as a **doublet of doublets**, abbreviated `**dd**`.

The signal for H_a at 5.95 ppm is also a doublet of doublets, with coupling constants $^3J_{ac}$ = 17.4 Hz and $^3J_{ab}$ = 6.31 Hz.



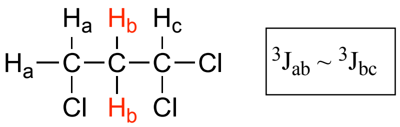
The signal for H_b at 5.64 ppm is split into a doublet by H_a, a *cis* coupling with $^3J_{ab} = 10.4$ Hz. Each of the resulting sub-peaks is split again by H_c, with the same *geminal* coupling constant $^2J_{bc} = 1.5$ Hz that we saw previously when we looked at the H_c signal. The overall result is again a doublet of doublets, this time with the two `sub-doublets` spaced slightly closer due to the smaller coupling constant for the *cis* interaction. Here is a blow-up of the actual H_b signal:



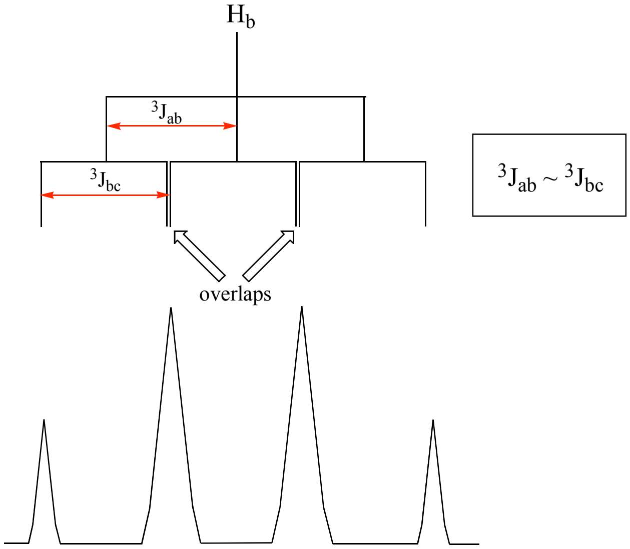
Exercise 5.9: Construct a splitting diagram for the H_b signal in the ^{1}H-NMR spectrum of methyl acrylate. Show the chemical shift value for each sub-peak, expressed in Hz (assume that the resonance frequency of TMS is exactly 300 MHz).

When constructing a splitting diagram to analyze complex coupling patterns, it is usually easier to show the larger splitting first, followed by the finer splitting (although the reverse would give the same end result).

When a proton is coupled to two different neighboring proton sets with identical or very close coupling constants, the splitting pattern that emerges often appears to follow the simple `*n* + 1 rule` of non-complex splitting. In the spectrum of 1,1,3-trichloropropane, for example, we would expect the signal for H_b to be split into a triplet by H_a, and again into doublets by H_c, resulting in a 'triplet of doublets'.

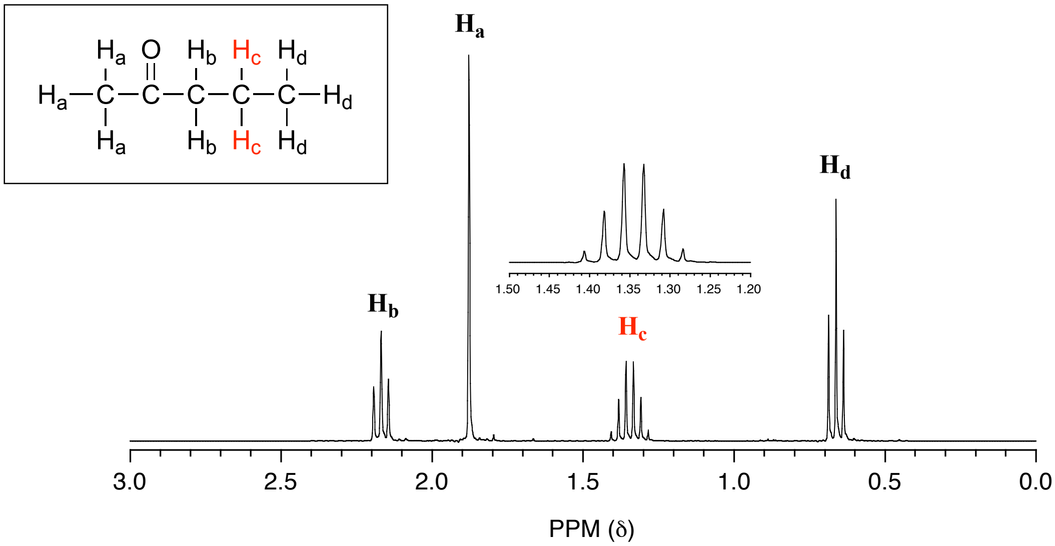


H_a and H_c are not equivalent (their chemical shifts are different), but it turns out that $^3J_{ab}$ is very close to $^3J_{bc}$. If we perform a splitting diagram analysis for H_b, we see that, due to the overlap of sub-peaks, the signal appears to be a quartet, and for all intents and purposes follows the n + 1 rule.

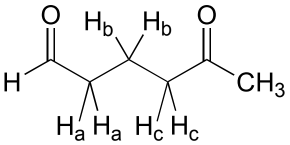


triplet of doublets becomes a quartet
when coupling constants are close

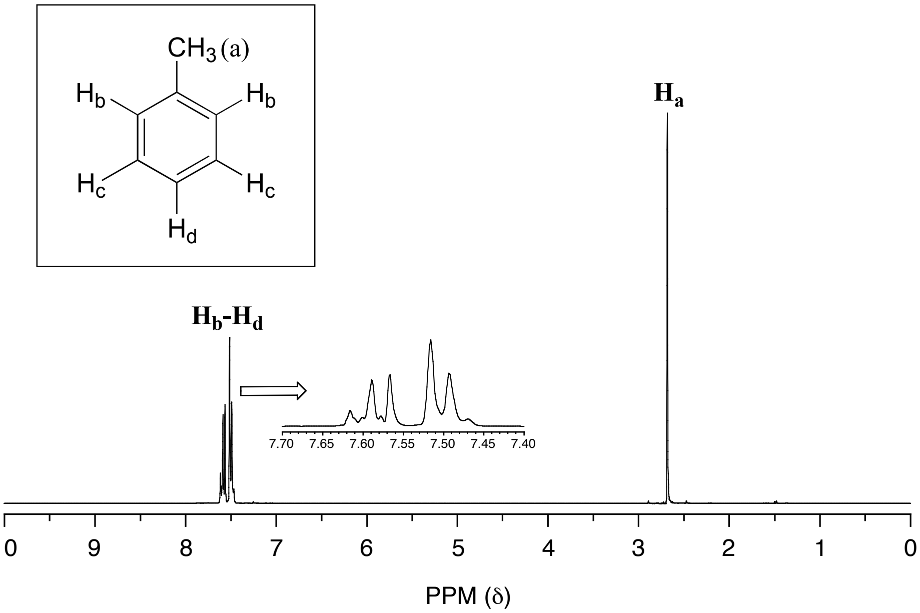
For similar reasons, the H_c peak in the spectrum of 2-pentanone appears as a sextet, split by the five combined H_b and H_d protons. Technically, this 'sextet' could be considered to be a 'triplet of quartets' with overlapping sub-peaks.



Exercise 5.10: What splitting pattern would you expect for the signal coresponding to H_b in the molecule below? Assume that $J_{ab} \sim J_{bc}$. Draw a splitting diagram for this signal, and determine the relative integration values of each subpeak.



In many cases, it is difficult to fully analyze a complex splitting pattern. In the spectrum of toluene, for example, if we consider only 3-bond coupling we would expect the signal for H_b to be a doublet, H_d a triplet, and H_c a triplet.



In practice, however, all three aromatic proton groups have very similar chemical shifts and their signals overlap substantially, making such detailed analysis difficult. In this case, we would refer to the aromatic part of the spectrum as a **multiplet**.

When we start trying to analyze complex splitting patterns in larger molecules, we gain an appreciation for why scientists are willing to pay large sums of money (hundreds of thousands of dollars) for higher-field NMR instruments. Quite simply, the stronger our magnet is, the more resolution we get in our spectrum. In a 100 MHz instrument (with a magnet of approximately 2.4 Tesla field strength), the 12 ppm frequency 'window' in which we can observe proton signals is 1200 Hz wide. In a 500 MHz (~12 Tesla) instrument, however, the window is 6000 Hz - five times wider. In this sense, NMR instruments are like digital cameras and HDTVs: better resolution means more information and clearer pictures (and higher price tags!)

Section 5.6: ^{13}C-NMR spectroscopy

The ^{12}C isotope of carbon - which accounts for up about 99% of the carbons in organic molecules - does not have a nuclear magnetic moment, and thus is NMR-inactive. Fortunately for organic chemists, however, the ^{13}C isotope, which accounts for most of the remaining 1% of carbon atoms in nature, has a magnetic moment just like protons. Most of what we have learned about ^{1}H-NMR spectroscopy also applies to ^{13}C-NMR, although there are several important differences.

5.6A: The basics of ^{13}C-NMR spectroscopy

The magnetic moment of a ^{13}C nucleus is much weaker than that of a proton, meaning that NMR signals from ^{13}C nuclei are inherently much weaker than proton signals. This, combined with the low natural abundance of ^{13}C, means that it is much more difficult to observe carbon signals: more sample is required, and often the data from hundreds of scans must be averaged in order to bring the signal-to-noise ratio down to acceptable levels. Unlike ^{1}H-NMR signals, the area under a ^{13}C-NMR signal cannot be used to determine the number of carbons to which it corresponds. This is because the signals for some types of carbons are inherently weaker than for other types – peaks corresponding to carbonyl carbons, for example, are much smaller than those for methyl or methylene (CH_2) peaks. Peak integration is generally not useful in ^{13}C-NMR spectroscopy, except when investigating molecules that have been enriched with ^{13}C isotope (see section 5.6B).

The resonance frequencies of ^{13}C nuclei are lower than those of protons in the same applied field - in a 7.05 Tesla instrument, protons resonate at about 300 MHz, while carbons resonate at about 75 MHz. This is fortunate, as it allows us to look at ^{13}C signals using a completely separate 'window' of radio frequencies. Just like in ^{1}H-NMR, the standard used in ^{13}C-NMR experiments to define the 0 ppm point is tetramethylsilane (TMS), although of course in ^{13}C-NMR it is the signal from the four equivalent *carbons* in TMS that serves as the standard. Chemical shifts for ^{13}C nuclei in organic molecules are spread out over a much wider range than for protons – up to 200 ppm for ^{13}C compared to 12 ppm for protons (see Table 3 for a list of typical ^{13}C-NMR chemical shifts). This is also fortunate, because it means that the signal from each carbon in a compound can almost always be seen as a distinct peak, without the overlapping that often plagues ^{1}H-NMR spectra. The chemical shift of a ^{13}C nucleus is influenced by essentially the same factors that influence a proton's chemical shift: bonds to electronegative atoms and diamagnetic anisotropy effects tend to shift signals downfield (higher resonance frequency). In addition, sp^{2} hybridization results in a large downfield shift. The ^{13}C-NMR signals for carbonyl carbons are generally the furthest downfield (170-220 ppm), due to both sp^{2} hybridization and to the double bond to oxygen.

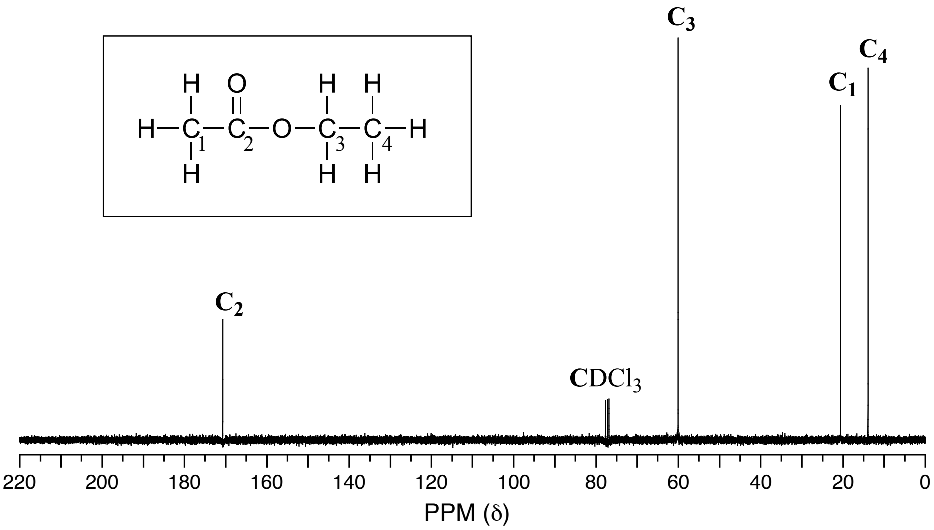
Exercise 5.11: How many sets of non-equivalent carbons are there in each of the molecules shown in exercise 5.1?

Exercise 5.12: How many sets of non-equivalent carbons are there in:

a) toluene
b) 2-pentanone
c) para-xylene
d) triclosan

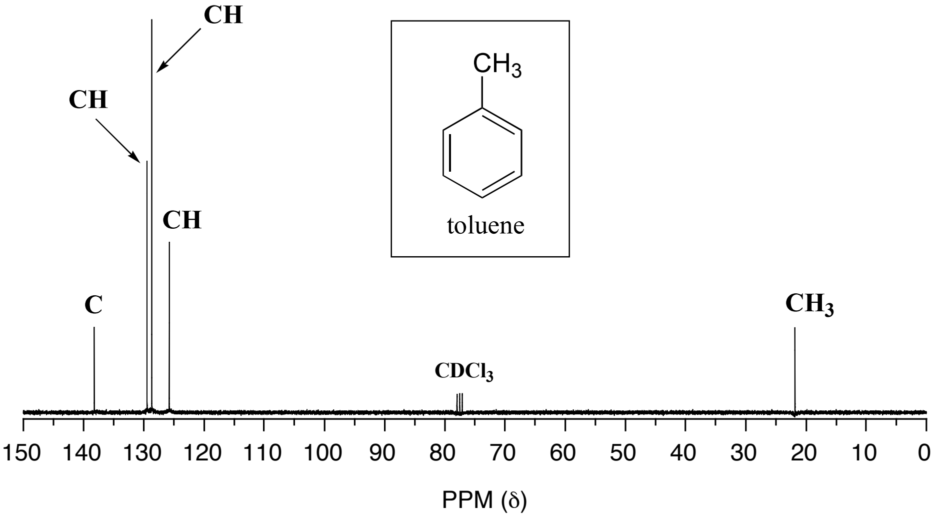
(all structures are shown earlier in this chapter)

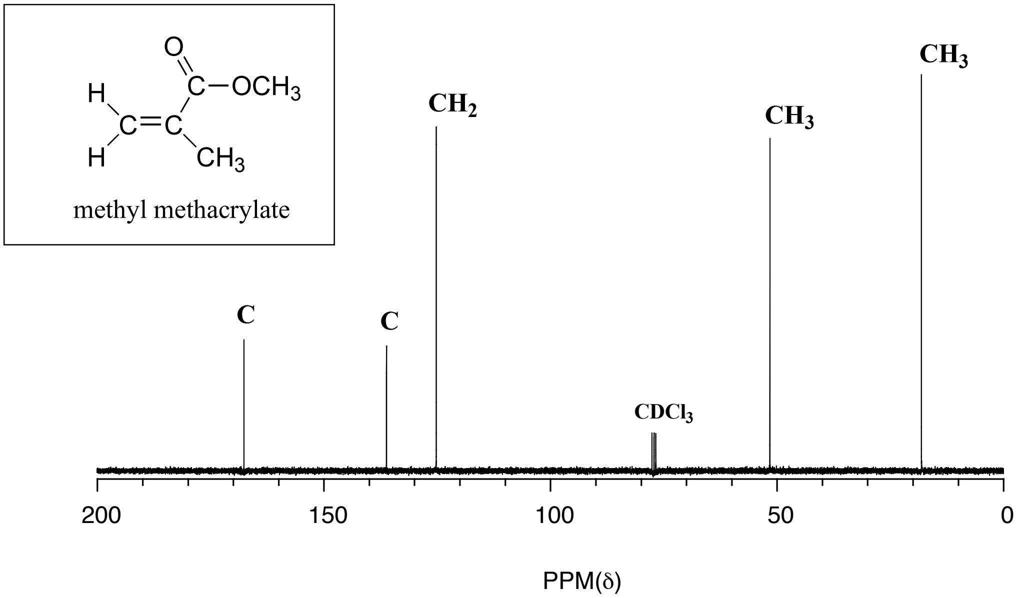
Because of the low natural abundance of ^{13}C nuclei, it is very unlikely to find two ^{13}C atoms near each other in the same molecule, and thus *we do not see spin-spin coupling between neighboring carbons in a ^{13}C-NMR spectrum*. There is, however, **heteronuclear coupling** between ^{13}C carbons and the hydrogens to which they are bound. Carbon-proton coupling constants are very large, on the order of 100 – 250 Hz. For clarity, chemists generally use a technique called **broadband decoupling**, which essentially 'turns off' C-H coupling, resulting in a spectrum in which all carbon signals are singlets. Below is the proton-decoupled ^{13}C-NMR spectrum of ethyl acetate, showing the expected four signals, one for each of the carbons.



While broadband decoupling results in a much simpler spectrum, useful information about the presence of neighboring protons is lost. However, another modern NMR technique called DEPT (Distortionless Enhancement by Polarization Transfer) allows us to determine how many hydrogens are bound to each carbon. For example, a DEPT experiment tells us that the signal at 171 ppm in the ethyl acetate spectrum is a quaternary carbon (no hydrogens bound, in this case a carbonyl carbon), that the 61 ppm signal is from a methylene (CH_2) carbon, and that the 21 ppm and 14 ppm signals are both methyl (CH_3) carbons. The details of the DEPT experiment are beyond the scope of this text, but DEPT information will often be provided along with ^{13}C spectral data in examples and problems.

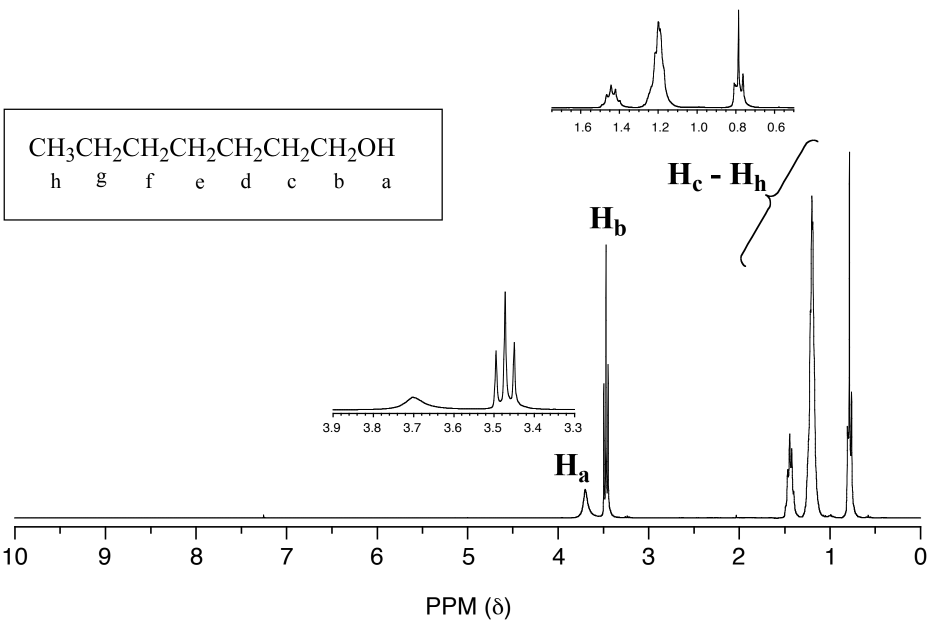
Below are two more examples of ^{13}C NMR spectra of simple organic molecules, along with DEPT information.



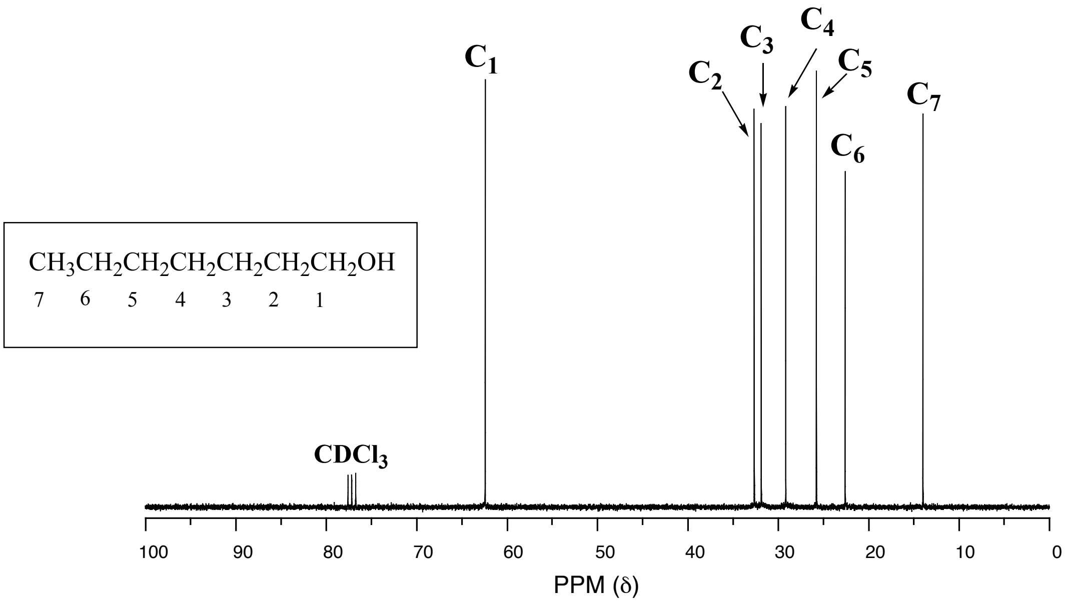


Exercise 5.13: Give peak assignments for the ^{13}C-NMR spectrum of methyl methacrylate, shown above.

One of the greatest advantages of ^{13}C-NMR compared to ^{1}H-NMR is the breadth of the spectrum - recall that carbons resonate from 0-220 ppm relative to the TMS standard, as opposed to only 0-12 ppm for protons. Because of this, ^{13}C signals rarely overlap, and we can almost always distinguish separate peaks for each carbon, even in a relatively large compound containing carbons in very similar environments. In the proton spectrum of 1-heptanol, for example, only the signals for the alcohol proton (H_a) and the two protons on the adjacent carbon (H_b) are easily analyzed. The other proton signals overlap, making analysis difficult.



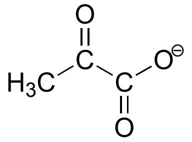
In the ^{13}C spectrum of the same molecule, however, we can easily distinguish each carbon signal, and we know from this data that our sample has seven non-equivalent carbons. (Notice also that, as we would expect, the chemical shifts of the carbons get progressively smaller as they get farther away from the deshielding oxygen.)



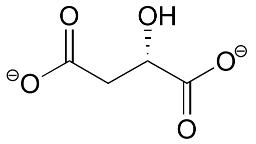
This property of ^{13}C-NMR makes it very helpful in the elucidation of larger, more complex structures.

Exercise 5.14: ^{13}C-NMR (and DEPT) data for some common biomolecules are shown below (data is from the Aldrich Library of ^{1}H and ^{13}C NMR). Match the NMR data to the correct structure, and make complete peak assignments.

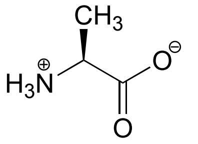
spectrum a: 168.10 ppm (C), 159.91 ppm (C), 144.05 ppm (CH), 95.79 ppm (CH)
spectrum b: 207.85 ppm (C), 172.69 ppm (C), 29.29 ppm (CH_3)
spectrum c: 178.54 ppm (C), 53.25 ppm (CH), 18.95 ppm (CH_3)
spectrum d: 183.81 ppm (C), 182. 63 ppm (C), 73.06 ppm (CH), 45.35 ppm (CH_2)



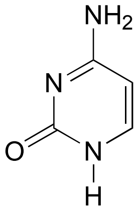
pyruvate



(S)-malate



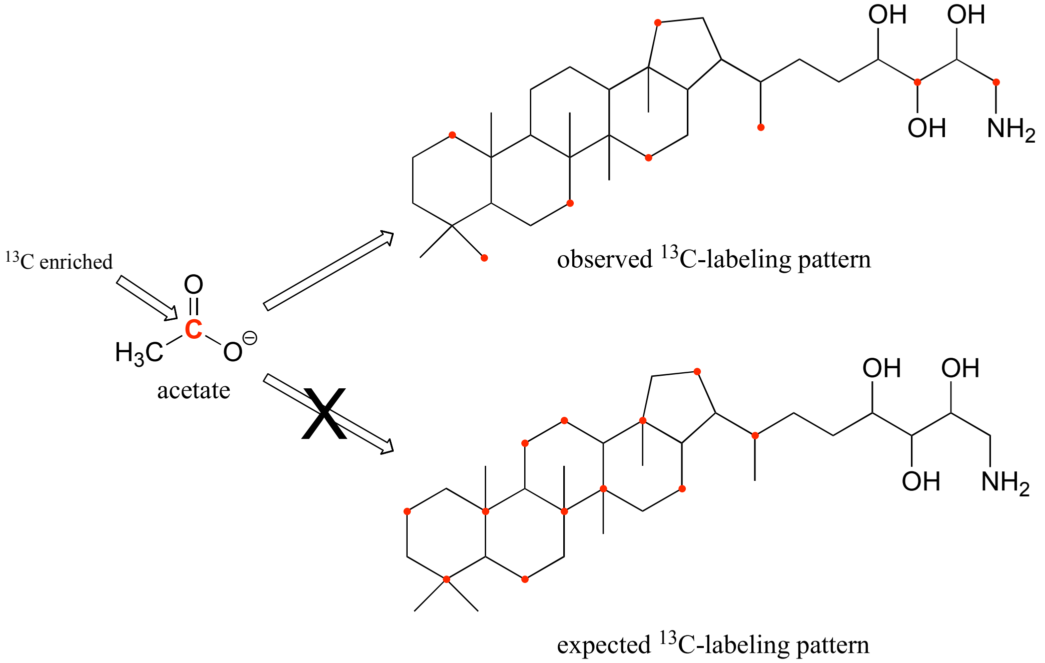
alanine



cytosine

5.6B: ^{13}C-NMR in isotopic labeling studies

Although only about 1 out of 100 carbon atoms in a naturally occurring organic molecule is the ^{13}C isotope, chemists are often able to synthesize molecules that are artificially enriched in ^{13}C at specific carbon positions, sometimes to the point where the ^{13}C isotope is incorporated at a given position in over half of the molecules in the sample. This can be very useful, especially in biochemical studies, because it allows us to 'label' one or more carbons in a small precursor molecule and then trace the presence of the ^{13}C label through a metabolic pathway all the way to a larger biomolecule product. For example, scientists were able to grow bacteria in a medium in which the only source of carbon was acetate enriched in ^{13}C at the C_1 (carbonyl) position. When they isolated a large molecule called amino-bacterio-hopanetriol (very similar in structure to cholesterol) from these bacteria and looked at its ^{13}C-NMR spectrum, they observed that the ^{13}C label from acetate had been incorporated at eight specific positions. They knew this because the ^{13}C-NMR signals for these carbons were much stronger compared to the same signals in a control (unlabeled) molecule.



This result was very surprising - the scientists had expected a completely different pattern of ^{13}C incorporation based on what they believed to be the metabolic pathway involved. This unexpected result led eventually to the discovery that bacteria synthesize these types of molecules in a very different way than yeasts, plants, and animals (*Eur. J. Biochem.* **1988**, *175*, 405). The newly discovered bacterial metabolic pathway is currently a key target for the development of new antibiotic and antimalaria drugs.

Section 5.7: Determining unknown structures

Now it is finally time to put together all that we have studied about NMR, IR, UV-Vis, and MS, and learn how to actually figure out the structure of an organic molecule 'from scratch' - starting, in other words, with nothing but raw experimental data. For this exercise, we will imagine that we have been given a vial containing a pure sample of an unknown organic compound, and that this compound to our knowledge has never before been synthesized, isolated, or characterized - we are the first to get our hands on it. Can we figure out its structure? (In reality, the examples that we use here are well-known compounds - but for now, just play along!).

Before we start analyzing spectroscopic data, we need one very important piece of information about our compound - its molecular formula. This can be determined through the combined use of mass spectrometry and **combustion analysis**. We will not go into the details of combustion analysis - for now, it is enough to know that this technique tells us the mass percent of each element in the compound. Because molecular oxygen is involved in the combustion reaction, oxygen in the sample is not measured directly - but we assume that if the mass percentages do not add up to 100%, the remainder is accounted for by oxygen.

When we obtain our unknown compound, one of the first things we will do is to send away a small quantity to an analytical company specializing in combustion analysis. They send us back a report stating that our compound is composed, by mass, of 52.0% carbon, 38.3% chlorine, and 9.7% hydrogen. This adds up to 100%, so our compound does not contain any oxygen atoms.

In order to determined the molecular formula of our compound from this data, we first need to know its molar mass. This piece of information, as you recall from chapter 4, we determine by looking at the 'molecular ion peak' in the mass spectrum of our compound. In this example, we find that our MS data shows a molecular ion peak at $m/z = 92$, giving us a molar mass of 92 g/mole (remember that in the MS experiment, charge (z) is almost always equal to 1, because we are looking at +1 cations).

So, one mole of our compound is 92g. How many moles of carbon atoms are in one mole of the compound? Simple: 52% of 92g is 47.8g. So in one mole of our compound, there is about 48 g of carbon, which means four moles of carbon. With similar calculations, we find that one mole of our compound contains nine hydrogens and one chlorine. Therefore our molecular formula is C_4H_9Cl. This formula has an index of hydrogen deficiency value (IHD, section 1.3D) of zero, meaning that our compound contains no multiple bonds or rings. We are off to a good start! Now, let's look at the data we get from NMR:

^{1}H-NMR:

3.38 ppm (2H, d, J = 6.13 Hz)
1.95 ppm (1H, m)
1.01 ppm (6H, d, J = 6.57)

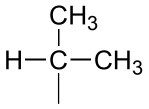
In case you can't decipher the above, here's some help. The spectrum contains three signals. The downfield signal, with a chemical shift of 3.38 ppm, integrates to two hydrogens, and is a doublet with a coupling constant of J = 6.13 Hz. The middle signal at 1.95 ppm is a multiplet (meaning we are not trying to analyze the complex splitting pattern), and integrates to one hydrogen. Finally, the upfield signal at 1.01 ppm integrates to six hydrogens, and is a doublet with J = 6.57 Hz). Now, on to the ^{13}C-NMR data:

^{13}C-NMR:

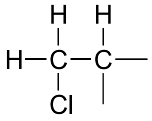
52.49 ppm (CH_2)
31.06 ppm (CH)
20.08 ppm (CH_3)

The process of piecing together an organic structure is very much like putting together a puzzle. In every case we start the same way, determining the molecular formula and the IHD value. After that, there is no set formula for success- what we need to do is figure out as much as we can about individual pieces of the molecule from the NMR (and often IR, MS, or UV-Vis) data, and write these down. Eventually, hopefully, we will be able to put these pieces together in a way that agrees with all of our empirical data. Let's give it a go.

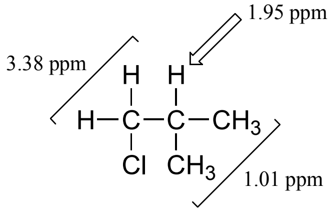
We see that there are only three signals in each NMR spectrum, but four carbons in the molecule, which tells us that two of the carbons are chemically equivalent. The fact that the signal at 1.01 ppm in the proton spectrum corresponds to *six* hydrogens strongly suggests that the molecule has two equivalent *methyl* (CH_3) groups. Because this signal is a doublet, there must be a CH carbon bound to each of these two methyl groups. Taken together, this suggests:



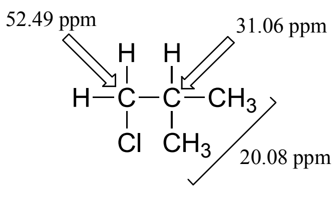
The ^{1}H-NMR signal at 3.38 ppm must be for hydrogens bound to the carbon which is in turn bound to the chlorine (we infer this because this signal is the furthest downfield in the spectrum, due to the deshielding effect of the electronegative chlorine). This signal is for two hydrogens, and is a doublet, meaning that these two hydrogens have a single hydrogen for a neighbor.



Putting all this together, we get 1-chloro-2-methylpropane.



^{1}H-NMR assignments



^{13}C-NMR assignments

Congratulations - you've solved your first organic structure!

Exercise 5.15:

Three compounds, both of which have the same combustion analysis results as 1-chloro-2-methylpropane, produce the following NMR data. Assign structures to the three compounds. Integration data has been omitted in order to make the problem more challenging.

Compound A:

^{1}H-NMR: 1.62 ppm (s)

^{13}C NMR: 67.14 ppm (C), 34. 47 ppm (CH_3)

Compound B:

^{1}H-NMR:

3.42 ppm (t)
1.68 ppm (m)
1.41 ppm (sextet)
0.92 ppm (t)

^{13}C-NMR:

44.74 ppm (CH_2)
34.84 ppm (CH_2)
20.18 (CH_2)
13.34 (CH_3)

Compound C:

^{1}H-NMR:

3.97 ppm (sextet)
1.71 (m)
1.50 (d)
1.02 (t)

^{13}C-NMR:

60.34 ppm (CH)
33.45 (CH_2)
24.94 (CH_3)
11.08 (CH_3)

Let's try another problem, this time incorporating IR information. The following data was obtained for a pure sample of an unknown organic compound:

Combustion analysis:

C: 85.7%
H: 6.67%

MS: Molecular ion at *m/z* = 210, base peak at *m/z* = 167.

^{1}H-NMR:

7.5-7.0 ppm (m, 10H)
5.10 ppm (s, 1H)
2.22 ppm (s, 3H)

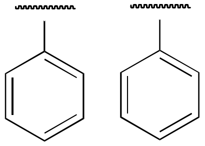
^{13}C-NMR:

206.2 (C)
138.4 (CH)
129.0 (CH)
128.7 (CH)
127.2 (CH)
65.0 (CH)
30.0 (CH_3)

IR: strong absorbance near 1720 cm^{-1}

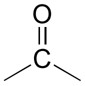
Our molecular weight is 210, and we can determine from combustion analysis that the molecular formula is $C_{15}H_{14}O$ (the mass percent of oxygen in the compound is assumed to be 100 - 85.7 - 6.67 = 7.6 %). This gives us IHD = 9.

Because we have ten protons with signals in the aromatic region (7.5-7.0 ppm), we are probably dealing with two phenyl groups, each with one substituted carbon.

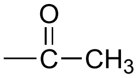


This accounts for 12 carbons, 10 hydrogens, and 8 IHD units. Notice that the carbon spectrum has only six peaks - and only four peaks in the aromatic region! This indicates that the two phenyl groups are equivalent.

The IR spectrum has a characteristic carbonyl absorption band, so that accounts for the oxygen atom in the molecular formula, the one remaining IHD unit, and the ^{13}C-NMR signal at 206.2 ppm.

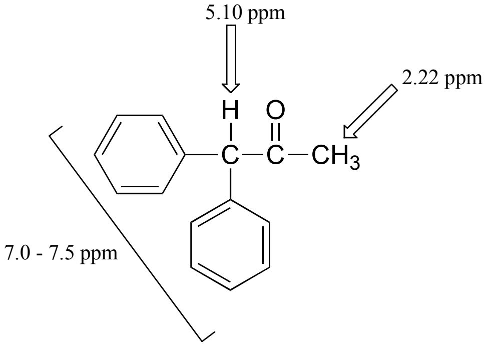


Now we only have two carbons and four hydrogens left to account for. The proton spectrum tells us we have a methyl group (the 2.22 ppm singlet) that is not split by neighboring protons - this chemical shift value is in the range of a methyl group adjacent to a carbonyl.

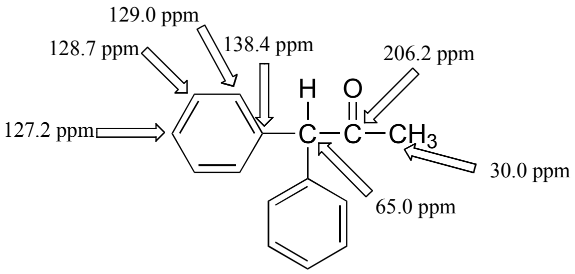


$$-\overset{\overset{\displaystyle O}{\|}}{C}-CH_3$$

Finally, there is one last proton at 5.10 ppm, also a singlet. Putting the puzzle together, the only possibility that fits is 1,1-diphenyl-2-propanone:

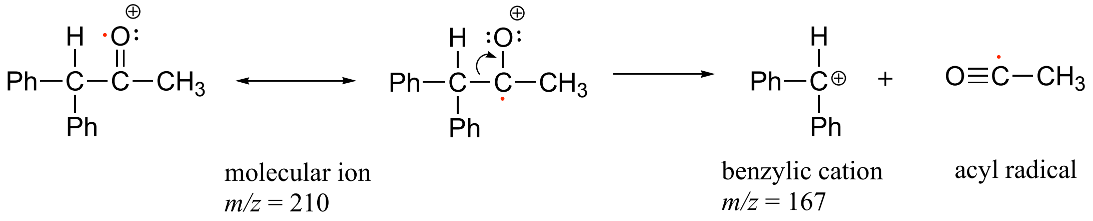


^{1}H-NMR assignments



^{13}C-NMR assignments

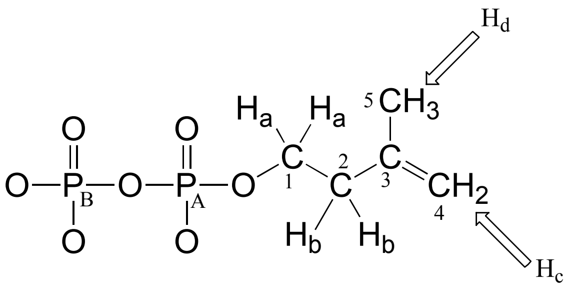
The base peak at *m/z* = 167 in the mass spectrum comes from loss of the acetyl group (see section 4.4B)



Exercise 5.16: Speculate as to why it is the fragment at *m/z* = 167 that forms the most abundant (base) peak in the mass spectrum of 1,1-diphenyl-2-propanone.

Section 5.8: NMR of phosphorylated molecules

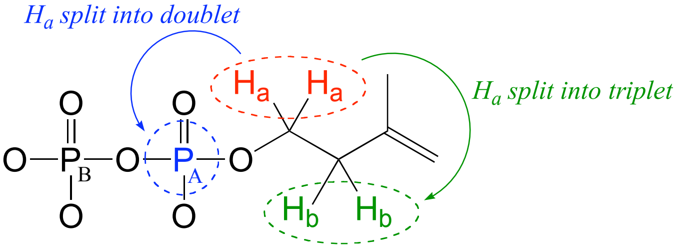
Because so many biological molecules contain phosphoryl groups, it is worthwhile to look at how scientists use NMR to determine the structure of these molecules. Consider the case of isopentenyl diphosphate, the building block molecule used by cells to make 'isoprenoid' compounds such as cholesterol (in many animals), or β-carotene (in some plants). NMR spectra of this molecule were taken in a D_2O solvent, buffered with ND_4OD (the deuterium equivalent of aqueous ammonium hydroxide, NH_4OH) (*J. Org. Chem*. **1986**, *51*, 4768).



^{1}H-NMR

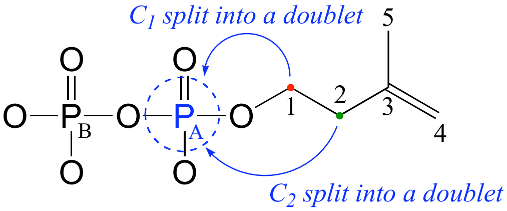
H_a: 4.05 ppm (td); $^{3}J_{Ha\text{-}Hb}$ = 6.6 Hz; $^{3}J_{Ha\text{-}PA}$ = 3.3 Hz.
H_b: 2.39 ppm (t) $^{3}J_{Ha\text{-}Hb}$ = 6.6 Hz
H_c: 4.86 ppm (s)
H_d: 1.77 ppm (s)

The ^{1}H-NMR signals for H_b, H_c, and H_d are what we would expect given our present knowledge. Why, though, is the signal for H_a split into a *triplet of doublets* (td)? First of all, as, expected, the two neighboring H_b protons split the H_a signal into a triplet, with ^{3}J = 6.6 Hz. Then, the signal is further split into doublets (^{3}J = 3.3 Hz) by P_A, the closer of the two phosphorus atoms.



Recall that the most abundant isotope of phosphorus is ^{31}P, which is NMR active - it has a magnetic dipole just like ^{13}C and ^{1}H, and thus will participate in spin-spin coupling interactions with nearby protons and ^{13}C nuclei.

In the ^{13}C spectrum, notice that the signals for both C_1 and C_2 are split into doublets by the magnetic field of P_A, through 2-bond and 3-bond coupling, respectively.

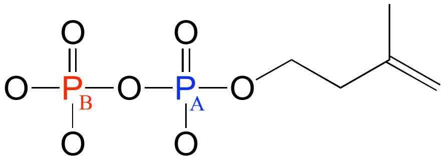


^{13}C-NMR (proton-decoupled)

C_1: 40.7 ppm (d); $^{2}J_{C1\text{-}PA}$ = 7.2 Hz
C_2: 67.0 ppm (d); $^{3}J_{C2\text{-}PA}$ = 4.0 Hz
C_3: 147.4 ppm
C_4: 114.6 ppm
C_5: 24.5 ppm

This splitting is observed even in the proton-decoupled spectrum - carbon-phosphorus coupling is *not* turned off.

Because ^{31}P is NMR-active, we can also, with the right instrumentation, *directly* observe the phosphorus NMR signals. In an NMR instrument with a 7.1 Tesla magnet, where protons resonate at 300 MHz and ^{13}C nuclei resonate at 75 MHz, phosphorus resonates at 32 MHz. In ^{31}P-NMR experiments, the reference standard used to determine the 0 ppm point is usually phosphoric acid (TMS, the standard for ^{1}H- and ^{13}C-NMR, doesn't have a phosphorus atom!). The ^{31}P-NMR spectrum of isopentenyl diphosphate has, as expected, two peaks, each of which is *upfield* of the phosphoric acid standard (negative chemical shifts!) and split into a doublet ($^{2}J_{P\text{-}P}$ = 20 Hz) due to coupling between the two phosphorus nuclei.



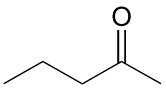
P_A: -11.03 ppm (d)
P_B: -7.23 ppm (d)

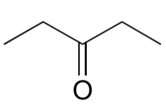
$^{2}J_{PA\text{-}PB}$ = 20 Hz

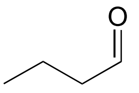
Notice that although the C_1 and C_2 signals were split by P_A in our ^{13}C-NMR spectrum, in the ^{31}P-NMR spectrum the P_A signal is *not* split by the nearby carbons - this is because, of course, both of these carbons are the NMR-inactive ^{12}C isotope in 99 out of 100 molecules. In addition, P-H splitting is not observed in this ^{31}P spectrum, because proton decoupling is in effect.

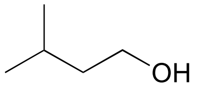
Chapter 5 problems

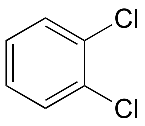
P5.1: For each molecule, predict the number of signals in the ^{1}H-NMR and the ^{13}C-NMR spectra (do not count split peaks - eg. a quartet counts as only one signal). Assume that diastereotopic groups are non-equivalent.

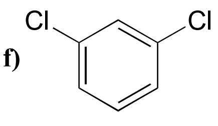
a)


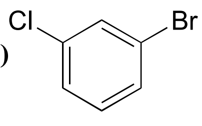
b)


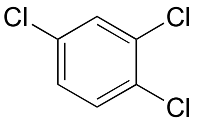
c)


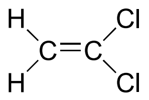
d)


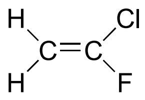
e)


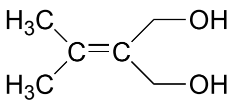
f)


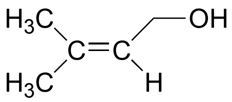
g)


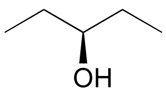
h)


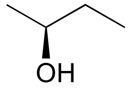
i)


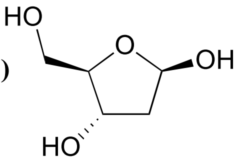
j)


k)


l)


m)


n)


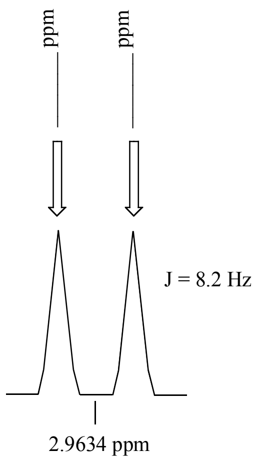
o)


P5.2: For each of the 20 common amino acids, predict the number of signals in the proton-decoupled ^{13}C-NMR spectrum.

P5.3: Calculate the chemical shift value (expressed in Hz, to one decimal place) of each sub-peak on the ^{1}H-NMR doublet signal below. Do this for:

a) a spectrum obtained on a 300 MHz instrument

b) a spectrum obtained on a 100 MHz instrument

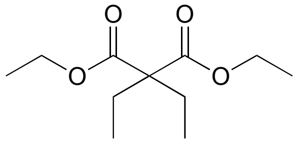


P5.4: Consider a quartet signal in an ^{1}H-NMR spectrum obtained on a 300 MHz instrument. The chemical shift is recorded as 1.7562 ppm, and the coupling constant is J = 7.6 Hz. What is the chemical shift, expressed to the nearest 0.1 Hz, of the furthest downfield sub-peak in the quartet? What is the resonance frequency (again expressed in Hz) of this sub-peak?)

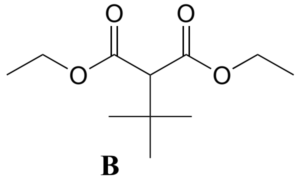
P5.5: One easily recognizable splitting pattern for the aromatic proton signals from disubstituted benzene structures is a pair of doublets. Does this pattern indicate *ortho*, *meta*, or *para* substitution?

P5.6: Match spectra below to their corresponding structures A-F.

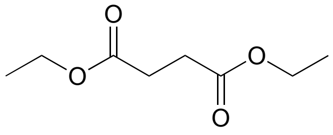
Structures:



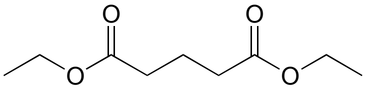
A



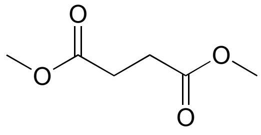
B



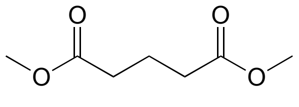
C



D



E



F

Spectrum 1:

δ (ppm)	splitting	integration
4.13	q	2
2.45	t	2
1.94	quintet	1
1.27	t	3

Spectrum 2:

δ (ppm)	splitting	integration
3.68	s	3
2.99	t	2
1.95	quintet	1

Spectrum 3:

δ (ppm)	splitting	integration
4.14	q	1
2.62	s	1
1.26	t	1.5

Spectrum 4:

δ (ppm)	splitting	integration
4.14	q	4
3.22	s	1
1.27	t	6
1.13	s	9

Spectrum 5

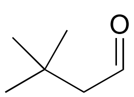
d (ppm)	splitting	integration
4.18	q	1
1.92	q	1
1.23	t	1.5
0.81	t	1.5

Spectrum 6:

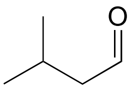
d (ppm)	splitting	integration
3.69	s	1.5
2.63	s	1

P5.7: Match spectra 7-12 below to their corresponding structures G-L .

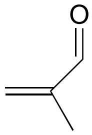
Structures:



G

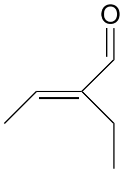


H

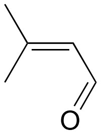


(J~0 for geminal vinylic protons)

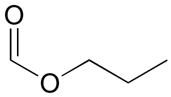
I



J



K



L

Spectrum 7:

δ (ppm)	**splitting**	**integration**
9.96	d	1
5.88	d	1
2.17	s	3
1.98	s	3

Spectrum 8:

δ (ppm)	**splitting**	**integration**
9.36	s	1
6.55	q	1
2.26	q	2
1.99	d	3
0.96	t	3

Spectrum 9:

δ (ppm)	splitting	integration
9.57	s	1
6.30	s	1
6.00	s	1
1.84	s	3

Spectrum 10:

δ (ppm)	splitting	integration
9.83	t	1
2.27	d	2
1.07	s	9

Spectrum 11:

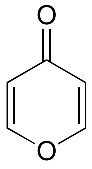
δ (ppm)	splitting	integration
9.75	t	1
2.30	dd	2
2.21	m	1
0.98	d	6

Spectrum 12:

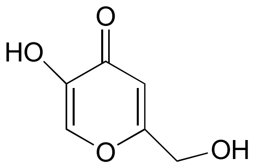
δ (ppm)	splitting	integration
8.08	s	1
4.13	t	2
1.70	m	2
0.96	t	3

P5.8: Match the ^{1}H-NMR spectra 13-18 below to their corresponding structures M-R .

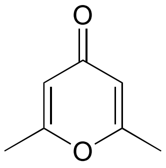
Structures:



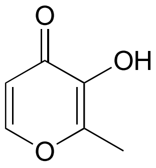
M



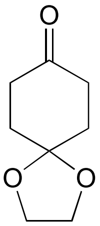
N



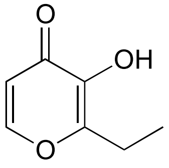
O



P



Q



R

Spectrum 13:

δ (ppm)	splitting	integration
8.15	d	1
6.33	d	1

Spectrum 14:

δ (ppm)	splitting	integration
6.05	s	1
2.24	s	3

Spectrum 15:

δ (ppm)	splitting	integration
8.57	s (b)	1
7.89	d	1
6.30	d	1
2.28	s	3

Spectrum 16:

δ (ppm)	splitting	integration
9.05	s (b)	1
8.03	s	1
6.34	s	1
5.68	s (b)	1
4.31	s	2

Spectrum 17:

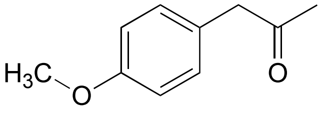
δ (ppm)	splitting	integration
7.76	d	1
7.57	s (b)	1
6.44	d	1
2.78	q	2
1.25	t	3

Spectrum 18:

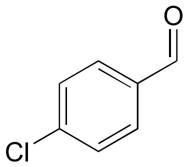
δ (ppm)	splitting	integration
4.03	s	1
2.51	t	1
2.02	t	1

P5.9: Match the ^{1}H-NMR spectra 19-24 below to their corresponding structures S-X.

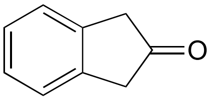
Structures:



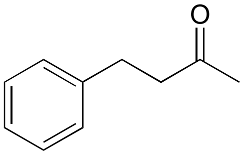
S



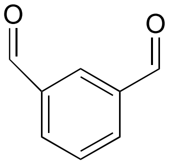
T



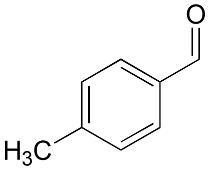
U



V



W



X

Spectrum 19:

δ (ppm)	splitting	integration
9.94	s	1
7.77	d	2
7.31	d	2
2.43	s	3

Spectrum 20:

δ (ppm)	splitting	integration
10.14	s	2
8.38	s	1
8.17	d	2
7.75	t	1

Spectrum 21:

δ (ppm)	splitting	integration
9.98	s	1
7.81	d	2
7.50	d	2

Spectrum 22:

δ (ppm)	**splitting**	**integration**
7.15-7.29	m	2.5
2.86	t	1
2.73	t	1
2.12	s	1.5

Spectrum 23:

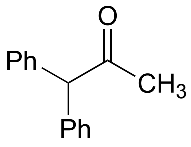
δ (ppm)	**splitting**	**integration**
7.10	d	1
6.86	d	1
3.78	s	1.5
3.61	s	1
2.12	s	1.5

Spectrum 24:

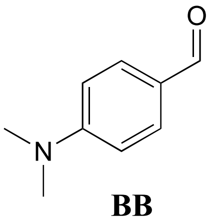
δ (ppm)	**splitting**	**integration**
7.23-7.30	m	1
3.53	s	1

P5.10: Match the ^{1}H-NMR spectra 25-30 below to their corresponding structures AA-FF.

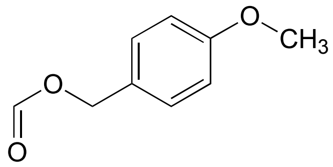
Structures:



AA

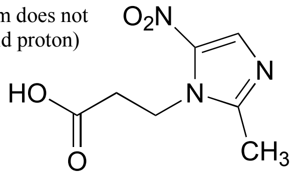


BB

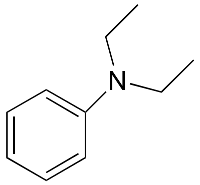


CC

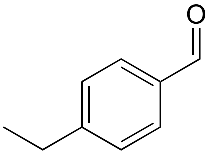
(spectrum does not show acid proton)



DD



EE



FF

Spectrum 25:

δ (ppm)	splitting	integration
9.96	s	1
7.79	d	2
7.33	d	2
2.72	q	2
1.24	t	3

Spectrum 26:

δ (ppm)	splitting	integration
9.73	s	1
7.71	d	2
6.68	d	2
3.06	s	6

Spectrum 27:

δ (ppm)	splitting	integration
7.20-7.35	m	10
5.12	s	1
2.22	s	3

Spectrum 28:

δ (ppm)	splitting	integration
8.08	s	1
7.29	d	2
6.87	d	2
5.11	s	2
3.78	s	3

Spectrum 29:

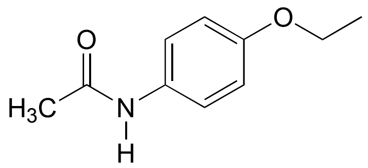
δ (ppm)	splitting	integration
7.18	d	1
6.65	m	1.5
3.2	q	2
1.13	t	3

Spectrum 30:

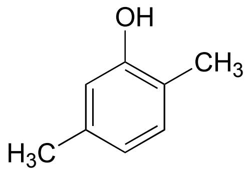
δ (ppm)	splitting	integration
8.32	s	1
4.19	t	2
2.83	t	2
2.40	s	3

P5.11: Match the ^{1}H-NMR spectra 31-36 below to their corresponding structures GG-LL

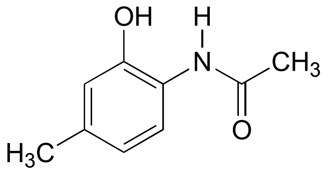
Structures:



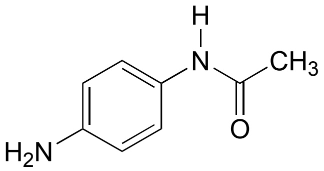
GG



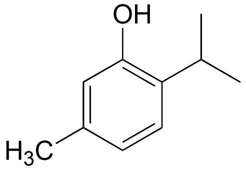
HH



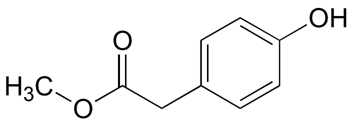
II



JJ



KK



LL

Spectrum 31:

δ (ppm)	**splitting**	**integration**
6.98	d	1
6.64	d	1
6.54	s	1
4.95	s	1
2.23	s	3
2.17	s	3

Spectrum 32:

δ (ppm)	**splitting**	**integration**
7.08	d	1
6.72	d	1
6.53	s	1
4.81	s	1
3.15	7-tet	1
2.24	s	3
1.22	d	6

Spectrum 33:

δ (ppm)	**splitting**	**integration**
7.08	d	2
6.71	d	2
6.54	s	1
3.69	s	3
3.54	s	2

Spectrum 34:

δ (ppm)	**splitting**	**integration**
9.63	s	1
7.45	d	2
6.77	d	2
3.95	q	2
2.05	s	3
1.33	t	3

Spectrum 35:

δ (ppm)	**splitting**	**integration**
9.49	s	1
7.20	d	2
6.49	d	2
4.82	s	2
1.963	s	3

Spectrum 36:

δ (ppm)	**splitting**	**integration**
9.58	s(b)	1
9.31	s	1
7.36	d	1
6.67	s	1
6.55	d	1
2.21	s	3
2.11	s	3

P5.12: Use the NMR data given to deduce structures.

a) Molecular formula: C_5H_8O

^{1}H-NMR:

δ (ppm)	**splitting**	**integration**
9.56	s	1
6.25	d (J ~ 1 Hz)	1
5.99	d (J ~ 1 Hz)	1
2.27	q	2
1.18	t	3

^{13}C-NMR

δ (ppm)	**DEPT**
194.60	CH
151.77	C
132.99	CH_2
20.91	CH_2
11.92	CH_3

b) Molecular formula: $C_7H_{14}O_2$

^{1}H-NMR:

δ (ppm)	**splitting**	**integration**
3.85	d	2
2.32	q	2
1.93	m	1
1.14	t	3
0.94	d	6

^{13}C-NMR

δ (ppm)	DEPT
174.47	C
70.41	CH_2
27.77	CH
27.64	CH_2
19.09	CH_3
9.21	CH_3

c) Molecular formula: $C_5H_{12}O$

^{1}H-NMR:

δ (ppm)	splitting	integration
3.38	s	2H
2.17	s	1H
0.91	s	9H

^{13}C-NMR

δ (ppm)	DEPT
73.35	CH_2
32.61	C
26.04	CH_3

d) Molecular formula: $C_{10}H_{12}O$

^{1}H-NMR:

δ (ppm)	splitting	integration
7.18-7.35	m	2.5
3.66	s	1
2.44	q	1
1.01	t	1.5

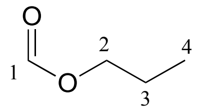
^{13}C-NMR

δ (ppm)	DEPT
208.79	C
134.43	C
129.31	CH
128.61	CH
126.86	CH
49.77	CH_2
35.16	CH_2
7.75	CH_3

P5.13:

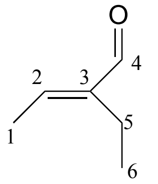
^{13}C-NMR data is given for the molecules shown below. Complete the peak assignment column of each NMR data table. For each structure, references are given to the structure in the Aldrich Library of NMR Spectra (volume - page).

a) (1-895C)



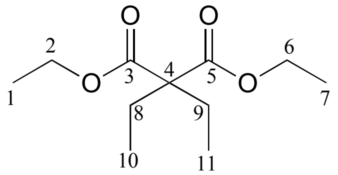
δ (ppm)	DEPT	carbon #(s)
161.12	CH	
65.54	CH_2	
21.98	CH_2	
10.31	CH_3	

b) (1-735A)



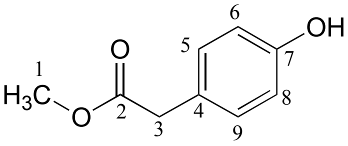
δ (ppm)	DEPT	carbon #(s)
194.72	C	
149.10	C	
146.33	CH	
16.93	CH_2	
14.47	CH_3	
12.93	CH_3	

c) (1-938)



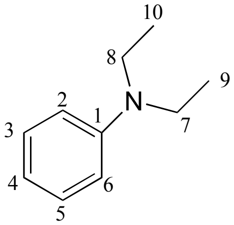
δ (ppm)	DEPT	carbon #(s)
171.76	C	
60.87	CH_2	
58.36	C	
24.66	CH_2	
14.14	CH_3	
8.35	CH_3	

d) (2-1212B)



δ (ppm)	DEPT	carbon #(s)
173.45	C	
155.01	C	
130.34	CH	
125.34	C	
115.56	CH	
52.27	CH_3	
40.27	CH_2	

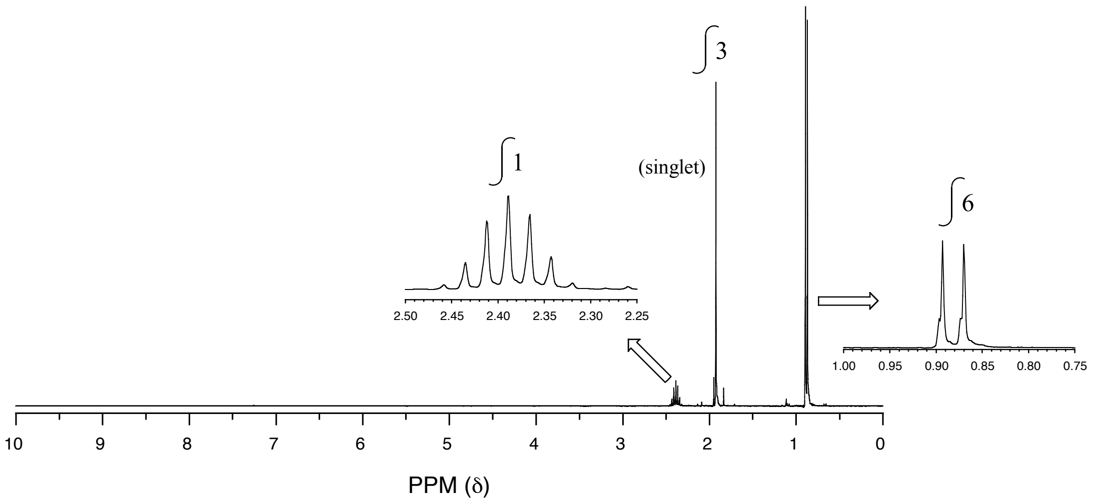
e) (2-455B)



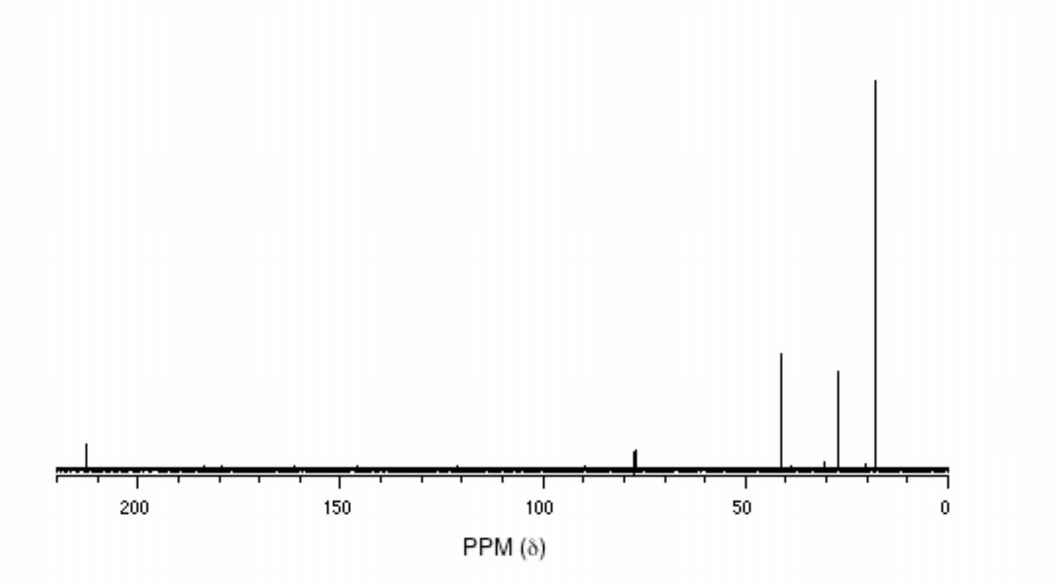
δ (ppm)	DEPT	carbon #(s)
147.79	C	
129.18	CH	
115.36	CH	
111.89	CH	
44.29	CH_2	
12.57	CH_3	

P5.14: You obtain the following data for an unknown sample. Deduce its structure.

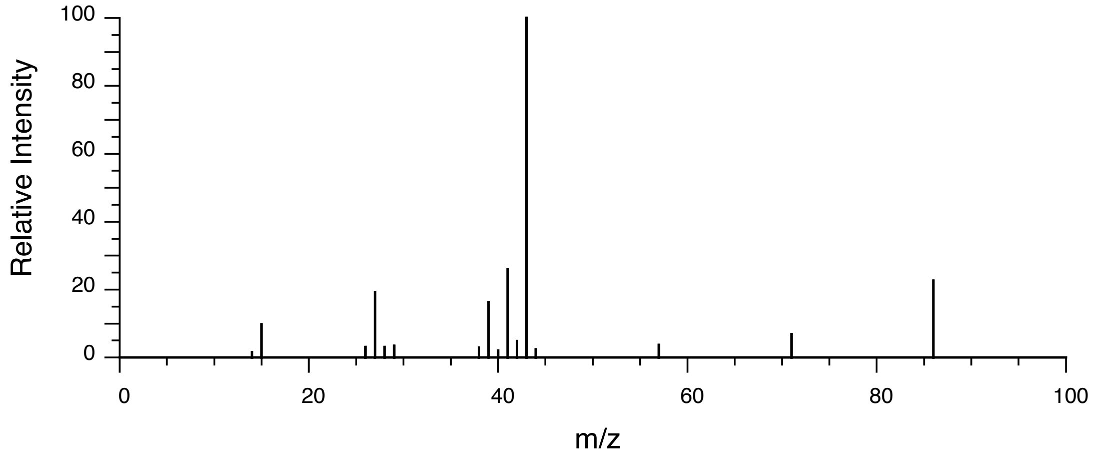
^{1}H-NMR:



^{13}C-NMR:



Mass Spectrometry:



P5.15: You take a ^{1}H-NMR spectrum of a sample that comes from a bottle of 1-bromopropane. However, you suspect that the bottle might be contaminated with 2-bromopropane. The NMR spectrum shows the following peaks:

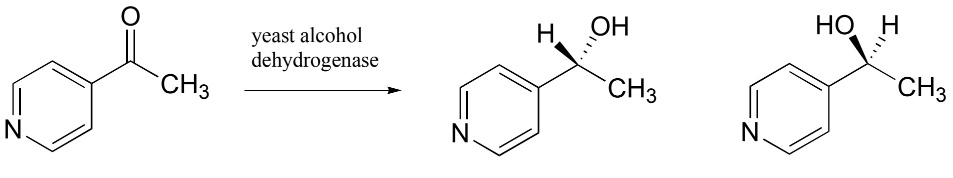
δ (ppm)	**splitting**	**Integration**
4.3	septet	0.0735
3.4	triplet	0.661
1.9	sextet	0.665
1.7	doublet	0.441
1.0	triplet	1.00

How badly is the bottle contaminated? Specifically, what percent of the molecules in the bottle are 2-bromopropane?

Challenge problems

C5.1: All of the ^{13}C-NMR spectra shown in this chapter include a signal due to $CDCl_3$, the solvent used in each case. Explain the splitting pattern for this signal.

C5.2: Researchers wanted to investigate a reaction which can be catalyzed by the enzyme alcohol dehydrogenase in yeast. They treated 4'-acylpyridine (1) with living yeast, and isolated the alcohol product(s) (some combination of 2A and 2B).

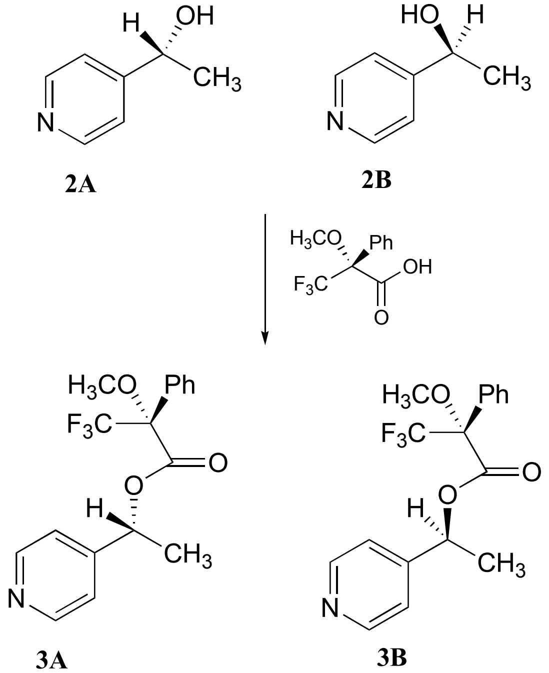


1 **2A** **2B**

a) Will the products 2A and 2B have identical or different ^{1}H-NMR spectra? Explain.

b) Suggest a ^{1}H-NMR experiment that could be used to determine what percent of starting material (1) got turned into product (2A and 2B).

c)With purified 2A/2B, the researchers carried out the subsequent reaction shown below to make 3A and 3B, known as 'Mosher's esters'. Do 3A and 3B have identical or different ^{1}H-NMR spectra? Explain.



2A **2B**

3A **3B**

d) Explain, very specifically, how the researchers could use ^{1}H-NMR to determine the relative amounts of 2A and 2B formed in the reaction catalyzed by yeast enzyme.

Chapter 6

Introduction to organic reactivity and catalysis

Introduction

Up to this point, we have been focusing on the *structure* of organic molecules: essentially, how these molecule are put together. The focus of the text will now shift to the study of their *reactivity*: what happens, in other words, when covalent bonds within a molecule break and new covalent bonds form, as molecule A is transformed into molecule B.

In your previous college general chemistry and high school chemistry courses (and perhaps in biology courses as well), you have no doubt seen many different examples of chemical reactions. Most likely, these reactions were depicted by chemical *equations*, showing the starting products (reactants) and the finished products connected by a 'reaction arrow'. In most cases, the structures of reactants and products were not considered - they were defined only by molecular formula and perhaps by physical state (solid, liquid, gas, or solution). The reaction below, showing the decomposition of dinitrogen pentoxide to nitrogen dioxide and oxygen, is a typical example of the 'equation' treatment of chemical reactivity.

$N_2O_2 \rightarrow 2NO_2 + \frac{1}{2}O_2$

More relevant to organic and biological chemistry, you probably at some point have also seen this chemical equation, showing how living cells break down glucose (a sugar) to form water and carbon dioxide:

$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$

This way of talking about chemical reactions is perfectly adequate in introductory chemistry classes, when fundamental chemical concepts like stoichiometry, thermodynamics, and kinetics are being explained for the first time. In organic chemistry, beginning with this chapter, we will go much further. We will certainly review the important fundamental concepts of thermodynamics and kinetics that you learned previously. But in our discussion of organic reactivity, we will bring our understanding of organic *structure* into the picture, and think about *how* reactions take place: which bonds break, which bonds form, *why* a particular bond breaks or forms, the order in which bond-breaking and bond-forming takes place, and the nature of any intermediate compounds that might form. Taken together, a description of a chemical reaction at this level is called a **reaction mechanism**. Beginning here, and continuing throughout the rest of the text, our main job will be to understand the mechanisms of the most important types of reactions undergone by organic molecules.

Section 6.1: A first look at reaction mechanisms

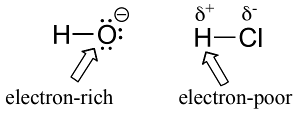
6.1A: An acid-base (proton transfer) reaction

For our first example of chemical reactivity, let's look at a very simple reaction that occurs between hydroxide ion and hydrochloric acid:

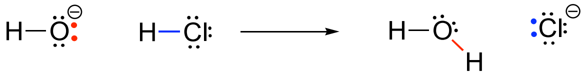
$HCl + OH^{-} \rightarrow H_2O + Cl^{-}$

This is an acid-base reaction: a proton is transferred from HCl, the acid, to hydroxide, the base. The product is water (the conjugate acid of hydroxide) and chloride ion (the conjugate base of HCl). You have undoubtedly seen this reaction before in general chemistry. Despite its simplicity (and despite the fact that the reactants and products are *inorganic* rather than organic), this reaction allows us to consider for the first time many of the fundamental ideas of organic chemistry that we will be exploring in various contexts throughout this text.

One very important key to understanding just about any reaction mechanism is the concept of **electron density**, and how it is connected to the **electron movement** (bond-breaking and bond-forming) that occurs in a reaction. The hydroxide ion – specifically, the electronegative oxygen atom in the hydroxide ion – has high electron density due to the polarity of the hydrogen-oxygen bond. The hydroxide oxygen is **electron-rich**.

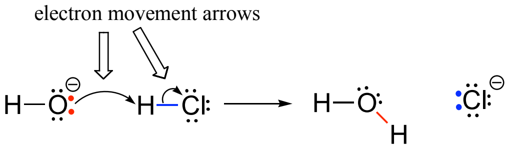


The hydrogen atom in HCl, on the other hand, has low electron density: it is **electron-poor**. As you might expect, something that is electron-rich is attracted to something that is electron-poor. As hydroxide and HCl move closer to each other, a lone pair of electrons on the electron-rich hydroxide oxygen is attracted by the electron-poor proton of HCl, and electron movement occurs towards the proton. The two electrons in the hydrogen-chlorine σ bond are repelled by this approaching hydroxide electron density, and therefore move even farther away from the proton and towards the chlorine nucleus. The consequence of all of this electron movement is that the hydrogen-chlorine bond is broken, as the two electrons from that bond completely break free from the 1s orbital of the hydrogen and become a lone pair in the 3p orbital of a chloride anion.



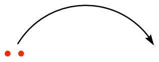
At the same time that the hydrogen-chlorine bond is breaking, a new σ bond forms between hydrogen and oxygen, containing the two electrons that previously were a lone pair on hydroxide. The result of this bond formation is, of course, a water molecule.

Previously (section 2.2) , we saw how curved arrows were used to depict 'imaginary' electron movement when drawing two or more resonance contributors for a single molecule or ion. These same curved arrows are used to show the very real electron movement that occurs in chemical reactions, where bonds are broken and new bonds are formed. The HCl + OH$^-$ reaction, for example, is depicted by drawing two curved arrows.

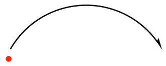


The first arrow originates at one of the lone pairs on the hydroxide oxygen and points to the 'H' symbol in the hydrogen bromide molecule, illustrating the 'attack' of the oxygen lone pair and subsequent formation of the new hydrogen-oxygen bond. The second curved arrow originates at the hydrogen-bromine bond and points to the 'Br' symbol, indicating that this bond is breaking - the two electrons are 'leaving' and becoming a lone pair on bromide ion.

It is very important to emphasize at this point that these curved, *two-barbed* arrows always represent the movement of *two* electrons. Most of this book will be devoted to the description of reaction mechanisms involving two-electron movement, so these full-headed arrows will become very familiar.



curved, two-barbed arrow:
two electron movement



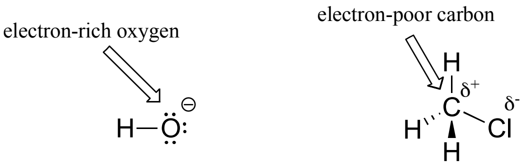
curved, single-barbed ('fish-hook') arrow:
single electron movement

In chapter 17, however, we will look at radical reaction mechanisms, where single-electron movement occurs. For these processes, a curved, single-barbed ('**fish-hook**') arrow will be used.

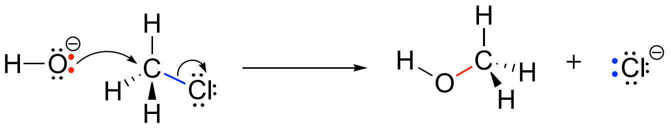
Exercise 6.1: Draw electron movement arrows to illustrate the acid-base reaction between acetic acid, CH_3COOH, and ammonia, NH_3. Draw out the full Lewis structures of reactants and products.

6.1B: A one-step nucleophilic substitution reaction (S_N2)

The reaction between hydroxide and HCl is a simple example of a Brønsted acid-base (proton transfer) reaction, and we will look at this reaction type in much more detail in chapter 7. For now, however, let's continue our introduction to the basic ideas of organic reactivity with a real organic reaction. Consider what might happen if a hydroxide ion encounters a chloromethane molecule instead of HCl. The hydroxide is still an electron-rich species, and thus might again be expected to act as a base and 'attack' a hydrogen. But in this case, the three hydrogens on the second reactant are not very electron-poor, as they are bound not to chlorine but to carbon, which is not very electronegative. However, there *is* a relatively electron-poor atom in chloromethane: the carbon.



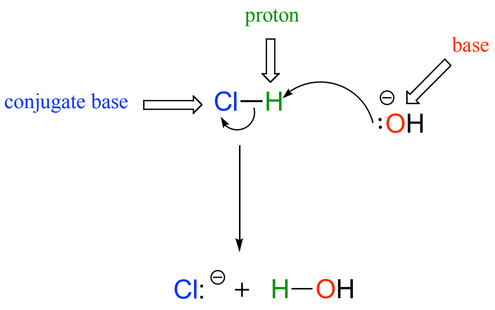
Because of the relative electronegativity of chlorine, the carbon-chlorine bond is polar. It stands to reason that a lone pair of electrons on the electron-rich hydroxide oxygen will be attracted to the electron-poor carbon.



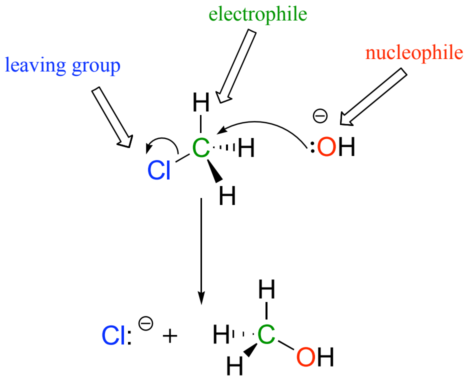
However, in order for a new bond to form between the hydroxide oxygen and the carbon, one of the bonds already on the carbon must break - otherwise, there will be five bonds to carbon and the octet rule will be violated. The C-Cl bond breaks as the new C-O bond forms, and the chlorine leaves along with its two electrons.

The reaction mechanism we see here is called a **nucleophilic substitution**, and is abbreviated S_N2. The 'substitution' term is easy to understand: just recognize how hydroxide substitutes for bromine as the fourth bond to the central carbon. The term 'nucleophilic' means 'nucleus-loving' and refers to the electron-rich species, the hydroxide oxygen. This oxygen is a **nucleophile**: it is attracted to the (positively-charged) nucleus of the central carbon atom, and 'attacks' with a lone pair of electrons to form a new covalent bond. The number '2' refers to the fact that this reaction has second order kinetics, a concept that we'll review shortly (section 6.2).

There are two more terms that come into play here, both of which you will see again and again as you continue to study organic reactions. Because the carbon atom in methyl bromide is electron-poor, it is attractive to anything that is electron rich - anything nucleophilic, in other words. Thus, the carbon is referred to in this context as an **electrophile**. The chlorine, because it leaves with its two electrons to become a chloride ion, is termed a **leaving group**. Notice that the three players in a nucleophilic substitution reaction - the nucleophile, the electrophile, and the leaving group - correspond conceptually to the three players in an acid-base reaction: the base, the acidic proton, and the conjugate base of the acid, respectively.



acid-base reaction

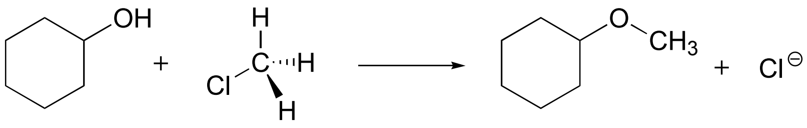


nucleophilic substitution

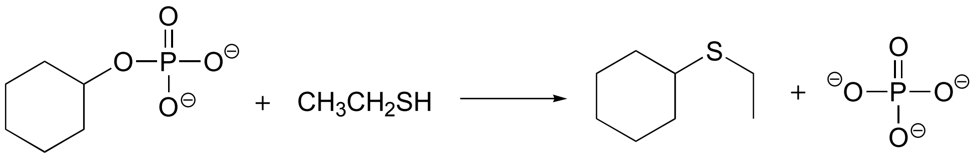
In many ways, the proton transfer process of an acid-base reaction can be thought of as simply a special kind of nucleophilic substitution reaction, one in which the electrophile is a hydrogen rather than a carbon.

Exercise 6.2: In each of the reactions below, identify the nucleophile, electrophile, and leaving group (assume in each case that a basic group is available to accept a hydrogen from the nucleophilic atom).

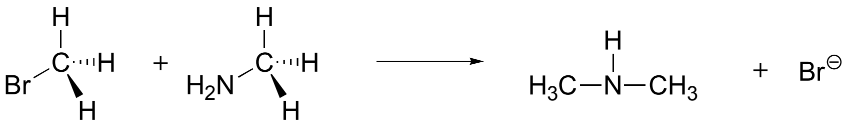
a)



b)

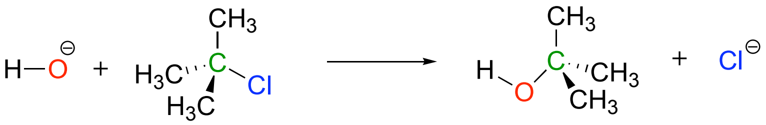


c)

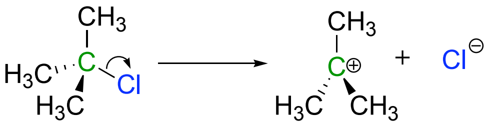


6.1C: A two-step nucleophilic substitution reaction (S_N1)

Reaction mechanisms describe not only the electron movement that occurs in a chemical reaction, but also the *order* in which bond-breaking and bond-forming events occur. Nucleophilic substitution reactions, for example, can occur by a second, alternative mechanism that is different from the mechanism above in terms of the order of events. A simple illustration is provided by the reaction of hydroxide with a tertiary alkyl chloride, such as 2-chloro-2-methyl propane.



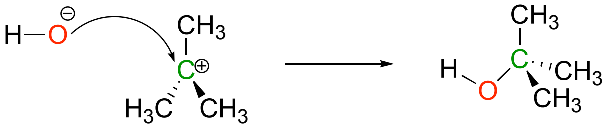
Unlike the chloromethane plus hydroxide reaction, in which the substitution process took place in a single, concerted step, this mechanism involves two separate steps. The leaving group, chloride anion, leaves first, *before* the hydroxide nucleophile approaches.



carbocation intermediate

What is left behind after the leaving group leaves is a carbocation: a planar, sp^2-hybridized carbon center with three bonds, an empty $2p_z$ orbital, and a full positive charge. In the language of organic mechanisms, this carbocation is referred to as a **reaction intermediate**.

A positively charged carbon is (obviously) very electron-poor, and thus the reactive intermediate is a powerful electrophile. It is quickly attacked by the hydroxide nucleophile to form the substitution product.



carbocation intermediate

We will see later (section 14.3) that other products are possible for this combination of reactants, but we will not worry about that for now.

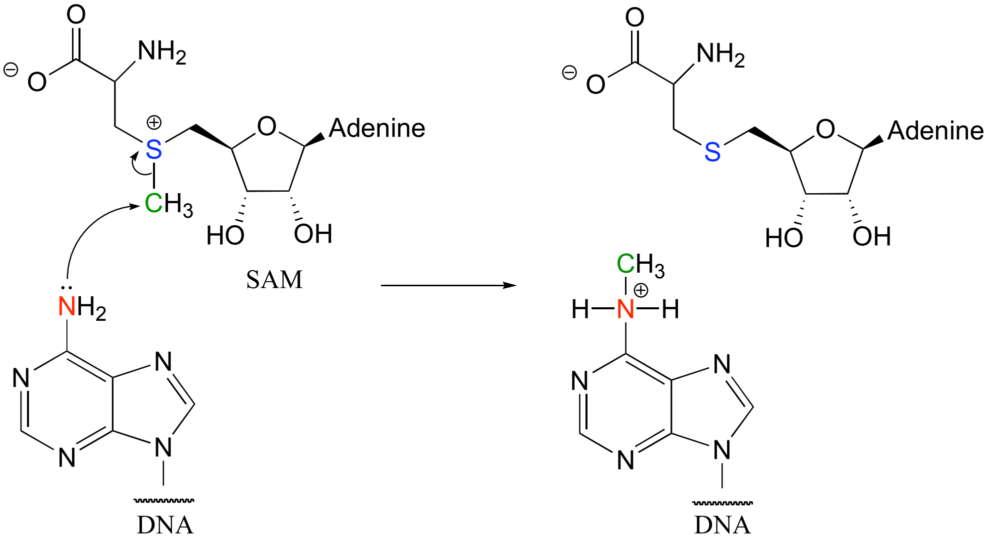
Note: In biological chemistry, the term 'intermediate' is also used to refer to compounds that are part of a metabolic pathway. In the general scheme below, compounds B, C, D, E, and F are all intermediate compounds in the metabolic pathway in which compound A is converted to compound G.

$$A \longrightarrow B \longrightarrow C \longrightarrow D \longrightarrow E \longrightarrow F \longrightarrow G$$

Pathway intermediates are often relatively stable compounds, whereas *reaction* intermediates (such as the carbocation species that plays a part in the two-step nucleophilic substitution) are short-lived, high energy species.

This mechanism is referred to by the abbreviation S_N1: a nucleophilic substitution with first order kinetics (again, the kinetic order part will be reviewed shortly in section 6.2).

Although nucleophilic substitutions at carbon are not terribly common in biochemistry, there are nevertheless some very important biological examples. One of these is DNA methylation. In the reaction below, the nucleophile is an amino nitrogen on adenosine (one of the four DNA building blocks). The electrophile is a methyl carbon on a molecule called S-adenosylmethionine (usually abbreviated ‘SAM’).

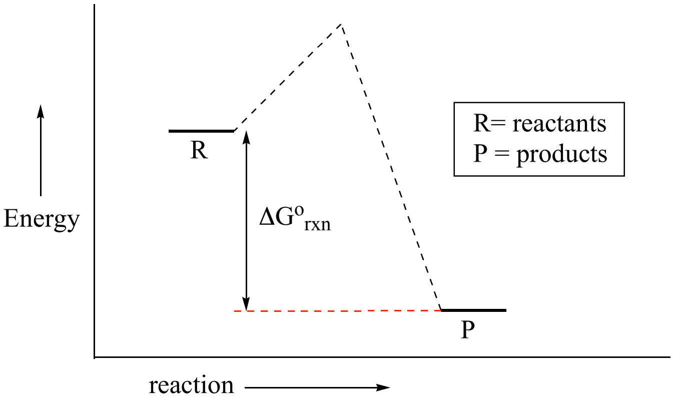


Notice that the leaving group in this reaction is a neutral sulfide, and that this is a single-step nucleophilic substitution (S_N2), like our chloromethane example.

We will have much more to say about nucleophilic substitutions, nucleophiles, electrophiles, and leaving groups in chapter 8, and we will learn *why* some substitutions occur in a single step and some occur in two steps with a carbocation intermediate. For now, however, we need to review the convention of energy diagrams and some of the basic concepts of thermodynamics and kinetics in order to continue our introduction to organic reactivity.

Section 6.2: Describing the thermodynamics and kinetics of chemical reactions - energy diagrams

You may recall from general chemistry that it is often convenient to describe chemical reactions with energy diagrams. In an energy diagram, the vertical axis represents the overall energy of the reactants, while the horizontal axis is the 'reaction coordinate', tracing from left to right the progress of the reaction from starting compounds to final products. The energy diagram for a typical S_N2 reaction might look like this:



Despite its apparent simplicity, this energy diagram conveys some very important ideas about the thermodynamics and kinetics of the reaction. Recall that when we talk about the **thermodynamics** of a reaction, we are concerned with the difference in energy between reactants and products, and whether a reaction is 'downhill' (exergonic, energy releasing) or 'uphill (endergonic, energy absorbing). When we talk about **kinetics**, on the other hand, we are concerned with the *rate* of the reaction, regardless of whether it is uphill or downhill thermodynamically.

First, let's review what this energy diagram tells us about the thermodynamics of the reaction illustrated by the energy diagram above. The energy level of the products is *lower* than that of the reactants. This tells us that the change in standard Gibbs Free Energy for the reaction (ΔG°_{rnx}) is negative. In other words, the reaction is exergonic, or 'downhill'. Recall that the ΔG°_{rnx} term encapsulates both ΔH°_{rnx}, the change in enthalpy (heat) and ΔS°_{rnx} , the change in entropy (disorder):

$\Delta G^\circ = \Delta H^\circ - T\, \Delta H^\circ$, where T is the temperature in Kelvin.

The standard Gibbs Free Energy change for a reaction can be related to the reaction's equilibrium constant (K_{eq}) by a simple equation:

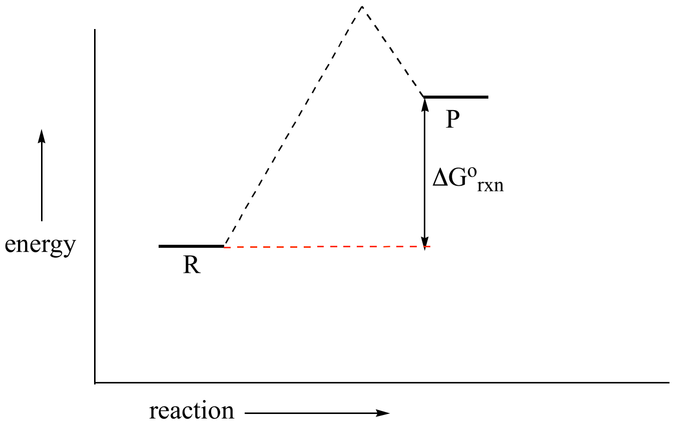
$\Delta G^\circ = -RT \ln K_{eq}$

where:

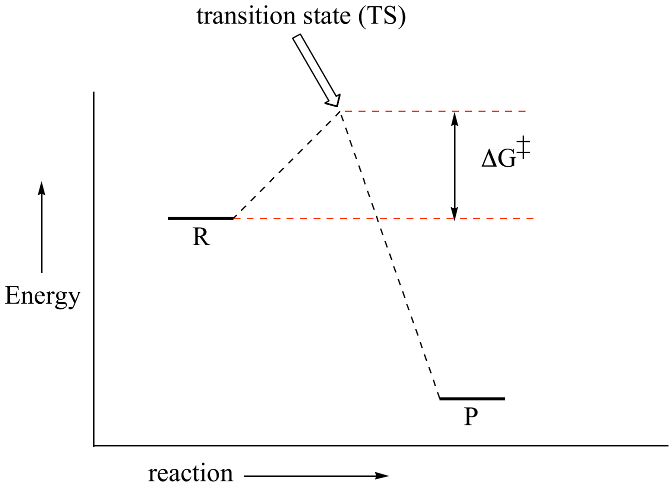
K_{eq} = [product] / [reactant] at equilibrium
$R = 8.314\ J{\cdot}K^{-1}{\cdot}mol^{-1}$ *or* $1.987\ cal{\cdot}\ K^{-1}{\cdot}mol^{-1}$
T = temperature in Kelvin (K)

If you do the math, you see that a negative value for ΔG°_{rnx} (an exergonic reaction) corresponds - as it should by intuition - to K_{eq} being greater than 1, an equilibrium constant which favors product formation.

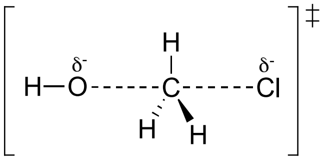
In a hypothetical endergonic (energy-absorbing) reaction the products would have a higher energy than reactants and thus ΔG°_{rnx} would be positive and K_{eq} would be less than 1, favoring reactants.



Now, let's move to kinetics. Look again at the energy diagram for exergonic reaction: although it is ‘downhill’ overall, it isn’t a straight downhill run.



First, an ‘energy barrier’ must be overcome to get to the product side. The height of this energy barrier, you may recall, is called the ‘**activation energy**’ ($\Delta G^‡$). The activation energy is what determines the kinetics of a reaction: the higher the energy hill, the slower the reaction. At the very top of the energy barrier, the reaction is at its **transition state** (TS), which is the point at which the bonds are in the process of breaking and forming. The transition state is an ‘**activated complex**’: a transient and dynamic state that, unlike more stable species, does not have any definable lifetime. It may help to imagine a transition state as being analogous to the exact moment that a baseball is struck by a bat. Transition states are drawn with dotted lines representing bonds that are in the process of breaking or forming, and the drawing is often enclosed by brackets. Here is a picture of a likely transition state for our simple S_N2 reaction between hydroxide and chloromethane:



The S_N2 reaction involves a collision between *two* molecules: for this reason, we say that it has **second order kinetics** (this is the source of the number ‘2’ in S_N2). The **rate expression** for this type of reaction is:

rate = k[reactant 1][reactant 2]

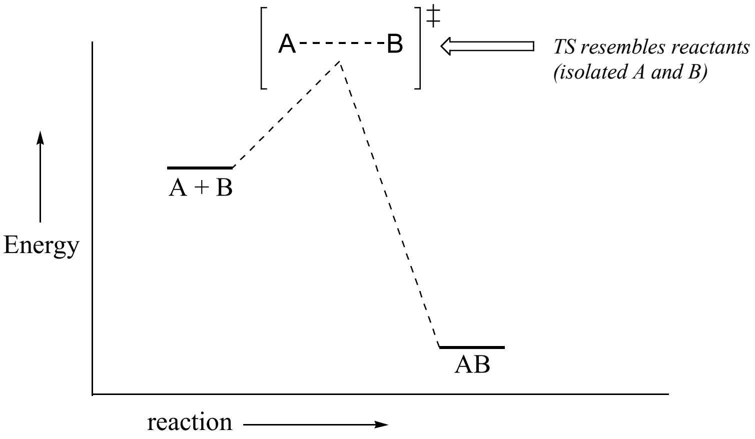
. . . which tells us that the rate of the reaction depends on the **rate constant** k as well as on the concentration of *both* reactants. The rate constant can be determined experimentally be measuring the rate of the reaction with different starting reactant concentrations. The rate constant depends on the activation energy, of course, but also on temperature: a higher temperature means a higher k and a faster reaction, all else being equal. This should make intuitive sense: when there is more heat energy in the system, more of the reactant molecules are able to get over the energy barrier.

Here is one more interesting and useful expression. Consider a simple reaction where the reactants are A and B, and the product is AB (this is referred to as a **condensation reaction**, because two molecules are coming together, or condensing). If we know the rate constant k for the forward reaction and the rate constant $k_{reverse}$ for the reverse reaction (where AB splits apart into A and B), we can simply take the quotient to find our equilibrium constant K_{eq}:

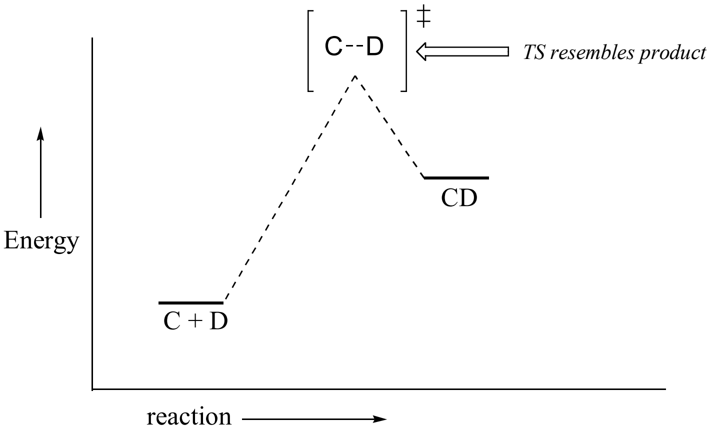
$$\mathrm{A + B} \rightleftharpoons \mathrm{AB} \qquad K_{eq} = \frac{[\mathrm{AB}]}{[\mathrm{A}]\,[\mathrm{B}]} = \frac{k_{forward}}{k_{reverse}}$$

This too should make some intuitive sense; if the forward rate constant is higher than the reverse rate constant, equilibrium should lie towards products.

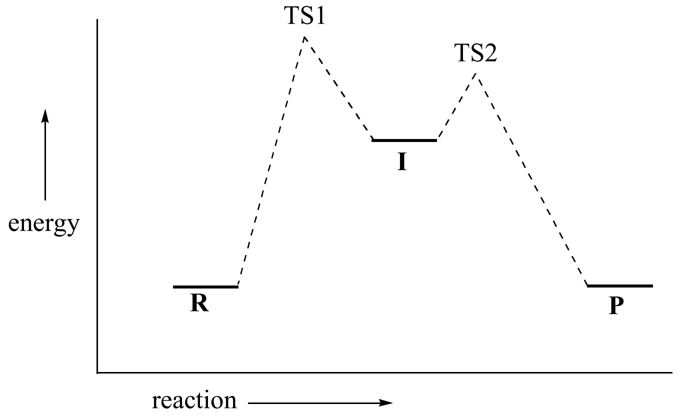
Now, back to transition states. Chemists are often very interested in trying to learn about what the transition state for a given reaction looks like, but addressing this question requires an indirect approach because the transition state itself cannot be observed. In order to gain some insight into what a particular transition state looks like, chemists often invoke the **Hammond postulate**, which states that *a transition state resembles the structure of the nearest stable species*. For an exergonic reaction, therefore, the transition state resembles the reactants more than it does the products.

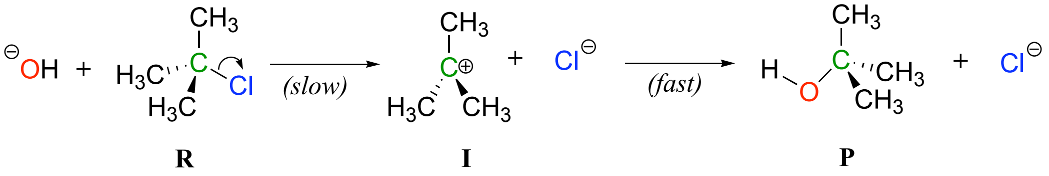


If we consider a hypothetical exergonic reaction between compounds A and B to form AB, the distance between A and B would be relatively large at the transition state, resembling the starting state where A and B are two isolated species. In the hypothetical endergonic reaction between C and D to form CD, however, the bond formation process would be much further along at the TS point, resembling the product.



Now, let's turn our attention to a two-step reaction mechanism, such as our S_N1 reaction between hydroxide and tert-butyl chloride. The energy diagram looks somewhat different:





Because there are two steps involved, there are also two transition states and two activation energies to consider, as well as the carbocation intermediate. The first, bond-breaking step from R to I can be depicted as a highly endergonic reaction, because the carbocation-chloride ion pair is significantly less stable (higher in energy) than the starting state. The transition state for this step, labeled TS1, can be thought of as that point at which the carbon-chlorine bond is halfway broken. The second step, attack on the carbocation electrophile by the hydroxide nucleophile and formation of the new carbon-oxygen bond, is a highly exergonic step that passes through a second, much lower transition state TS2. The intermediate (I) is thus depicted as an energy valley situated between TS1 and TS2. According to Hammond's postulate, both TS1 and TS2 in this reaction should resemble the intermediate I more than either the reactants or products, because in both cases the intermediate is the structure that is closest in energy. This idea – that an intermediate species resembles the nearby transition states - is a very important point to keep in mind, and we will revisit it several times throughout this text as we consider various organic reaction mechanisms.

There is one more very important idea expressed by the energy diagram that relates to the kinetics of the reaction. The activation energy for step 1 is much higher than the activation energy for step 2, meaning that step 1 is slower than step 2. This should make intuitive sense, because step 1 involves bond-breaking and separation of charge, while step 2 involves bond-forming and neutralization of charge. Because step 1 is slow and step 2 is fast, the overall rate of the reaction depends only on the rate of step 1. Step 1 is the '**rate-determining step**'; it is the bottleneck for the overall process. Anything that speeds up the rate of step 2 will *not* have any effect on the overall rate of the reaction.

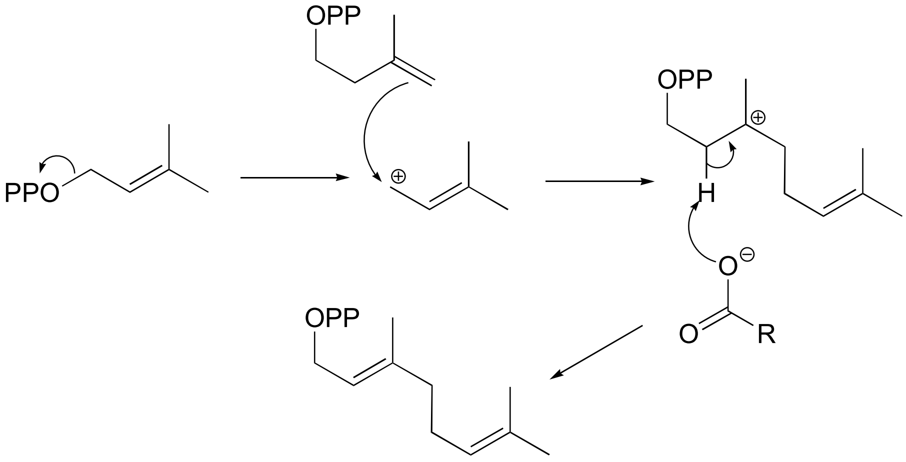
Notice also that the rate-determining step involves only *one* of the two reactants: *tert*-butyl chloride breaks apart in this step, while hydroxide is not involved at all (contrast this to the S_N2 reaction, where the single concerted step involves the collision of *two* species). The S_N1 reaction has **first order kinetics**, and the rate depends only on the rate constant *k* and the concentration of *tert*-butyl chloride:

rate = k[*tert*-butyl chloride]

What this means is that if we double the rate of *tert*-butyl chloride, the rate of the reaction will also double. Doubling the hydroxide ion concentration, however, will have no effect on the rate, because hydroxide only participates in the fast, non-rate-determining step.

When thinking about any chemical reaction, it is very important not to confuse the two fundamental chemical concepts of thermodynamics and kinetics. A ‘downhill’, energy-releasing reaction can have a low activation energy and be very fast (our HCl + OH^- example is a good case in point). A downhill reaction could also have a high activation energy and be very slow. Think of sugar: you know that sucrose is a high energy molecule and that you obtain lots of energy when you eat it and convert it to carbon dioxide and water, but you also know that sugar sitting in a bowl does not just spontaneously dissipate into carbon dioxide and water before your eyes. This process, although downhill in terms of thermodynamics, occurs *very* slowly in the absence of heat or catalysts.

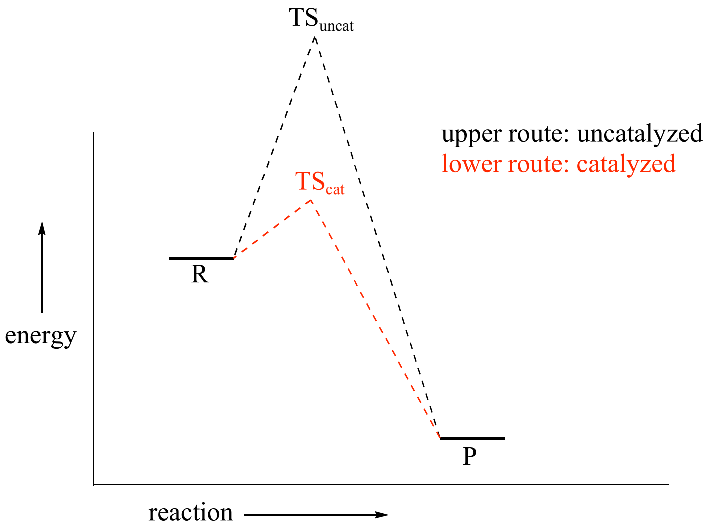
Exercise 6.3: Draw a two-dimensional representation of the transition state of each step in the reaction below (this is a reaction that you do not need to understand in detail now - we will study it in chapter 13).



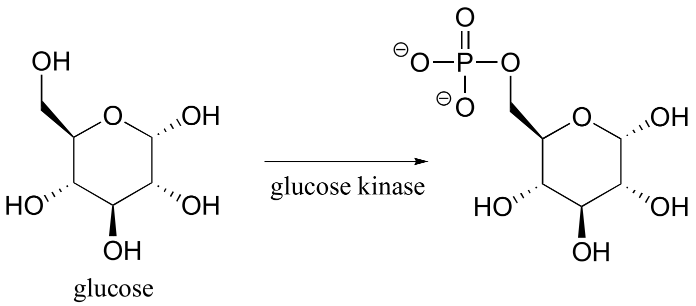
Section 6.3: Enzymatic catalysis - the basic ideas

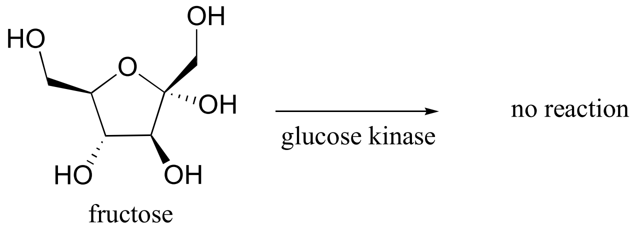
We come now to the subject of catalysis. Our hypothetical bowl of sugar is still stubbornly refusing to turn into carbon dioxide and water, even though by doing so it would reach a much more stable energy state. There are, in fact, two ways that we could speed up the process so as to avoid waiting several millennia for the reaction to reach completion. We could supply enough energy, in the form of heat from a flame, to push some of the sugar molecules over the high energy hill. Heat would be released from the resulting exothermic reaction, and this energy would push more molecules over their energy hills, and so on - the sugar would literally burn up.

A second way to make the reaction go faster is to employ a **catalyst**. You probably already know that a catalyst is an agent that causes a chemical reaction to go faster by lowering its activation energy.



How might you catalyze the conversion of sugar to carbon dioxide and water? It's not too hard – just eat the sugar, and let your digestive enzymes go to work catalyzing the many biochemical reactions involved in breaking it down. Enzymes are proteins, and are very effective catalysts. 'Very effective' in this context means very specific, and very fast. Most enzymes are very selective with respect to reactant molecules: they have evolved over millions of years to catalyze their specific reactions. An enzyme that attaches a phosphate group to glucose, for example, will not do anything at all to fructose (the details of these reactions are discussed in section 10.2B).





Glucose kinase is able to find and recognize glucose out of all of the other molecules floating around in the 'chemical soup' of a cell. A different enzyme, fructokinase, specifically catalyzes the phosphorylation of fructose.

We have already learned (section 3.9) that enzymes are very specific in terms of the stereochemistry of the reactions that they catalyze . Enzymes are also highly **regiospecific**, acting at only one specific part of a molecule. Notice that in the glucose kinase reaction above only one of the alcohol groups is phosphorylated.

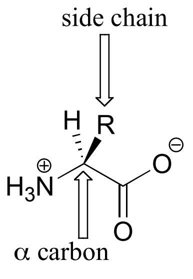
Finally, enzymes are capable of truly amazing rate acceleration. Typical enzymes will speed up a reaction by anywhere from a million to a billion times, and the most efficient enzyme currently known to scientists is believed to accelerate its reaction by a factor of about 10^{17} (see *Chemical and Engineering News*, March 13, 2000, p. 42 for an interesting discussion about this enzyme, orotidine monophosphate decarboxylase).

We will now begin an exploration of some of the basic ideas about how enzymes accomplish these amazing feats of catalysis, and these ideas will be revisited often throughout the rest of the text as we consider various examples of enzyme-catalyzed organic reactions. But in order to begin to understand how enzymes work, we will first need to learn (or review, as the case may be) a little bit about protein structure.

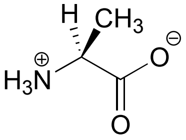
Section 6.4: Protein structure

6.4A: Amino acids and peptide bonds

Proteins are long chains of small 'building block' molecules called amino acids. There are twenty different amino acids generally found in proteins, all of which are based on the common structural framework shown below.



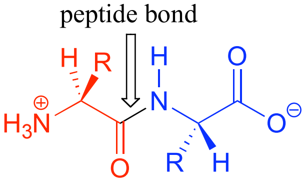
Amino acids contain an amino group on one side and a carboxylate group on the other, with a substituted carbon, referred to as the 'α-carbon', in the middle. It is the substituent on the α-carbon – designated 'R' in the general structure here, and often referred to as the amino acid 'side chain' – that distinguishes the 20 different amino acids. In all but one of the twenty, the α-carbon is a chiral center (in glycine, the 'R' group is simply a hydrogen, and thus the α-carbon is achiral). Virtually all of the amino acids found in proteins have the stereochemistry shown in the general figure above, and are referred to as L-amino acids ('L' stand for levorotary, the property of rotating plane polarized light in a counter-clockwise direction). Each amino acid is designated by both a three-letter and a single-letter abbreviation: the amino acid alanine, for example, is designated 'Ala' or just 'A'.



alanine
Ala
A

The side chains of amino acids contain a variety of functional groups (see table 5 in the tables section). Some like alanine have only alkyl functionality, while others have alcohol, thiol, sulfide, or amide groups. Four amino acids have aromatic groups on their side chains: phenylalanine has a phenyl group, tyrosine a phenol, tryptophan an indole, and histidine an imidazole. Many of the side chains are 'ionizable', capable of acting as acids or bases depending on their protonation state. This property makes them extremely important in enzyme catalysis. Aspartate and glutamate, for example, both have carboxylate groups and are deprotonated at pH 7. Lysine and arginine have amine and imino groups, respectively, and are protonated (and positively charged) at pH 7. Histidine, with its imidazole group, is mainly protonated and positively charged at pH 7. The phenol group of tyrosine is also capable of donating a proton. We will have more to say later about the pK_a values of the different side chains in chapter 7. As you progress in your study of organic chemistry, it will become very helpful to familiarize yourself with the structures of all 20 amino acids.

Proteins are formed when amino acids are linked together by 'peptide bonds' to form long chains (a peptide bond is really just a specific type of amide functional group). When they are part of a protein chain, individual amino acids are often referred to as **'residues'**. The figure below shows a 'dipeptide' - two amino acids linked together by a peptide bond.



Many protein chains are several hundred amino acids long. The 'backbone', or 'main chain' of a protein refers to the repeating nitrogen - α-carbon – carbonyl pattern: in other words, everything but the side chain. As we have discussed previously, the peptide bond has significant double-bond character due to resonance, and therefore there is a considerable barrier to rotation. Other σ bonds in proteins, however, have more rotational flexibility.

6.4B: Visualizing protein structure: X-ray crystallography

The key to the amazing versatility of proteins lies in their ability to fold up into very specific three dimensional structures. Much of what we know about the three-dimensional structure of proteins has been gained from a technique called **x-ray crystallography**. The details are beyond the scope of this text, but the basic idea is fairly easy to grasp. Essentially, patterns formed when x-rays are diffracted by a protein crystal allow for the generation of a three-dimensional picture of electron density. The electron density information can then be translated to the form of a map, on x, y, and z coordinates, of every non-hydrogen atom in the protein. Computer programs can use this data to generate convenient pictures of the protein.

Several of the figures that we'll be looking at in this section were generated using x-ray crystallography data retrieved from a database called the 'Protein Data Bank' (PDB), and visualized using a program called 'Jmol'. These are free, easy to use web-based resources - the URL addresses are:

Protein databank: http://www.rcsb.org/pdb/home/home.do
Jmol: http://molvis.sdsc.edu/fgij/index.htm

Figures in this section were derived from x-ray crystal structures of the following enzymes:

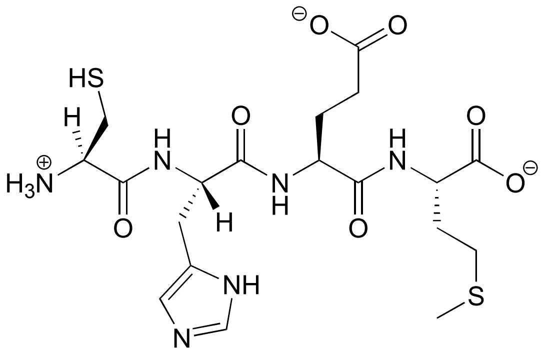
Fructose 1,6-bisphosphate aldolase (*Protein Science* **1999**, *8*, 291; pdb 4ALD)
Fructose 6-phosphate aldolase (*J. Mol. Biol*. **2002**, *319*, 161; pdb 1l6w)
DNA adenine methyltransferase (*Structure* **1998**, *6*, 1563; pdb 2DPM)
Human transaldolase (*FEBS Lett*. **2000**, *475*, 204; pdb 1FO5
Protein farnesyltransferase (*Biochemistry* **1998**, *37*, 9612; pdb 1FT2)

You can look at these structures, rotate them around, zoom in and out, and focus on different groups if you visit the Protein Data Bank and enter the 'pdb' code provided above.

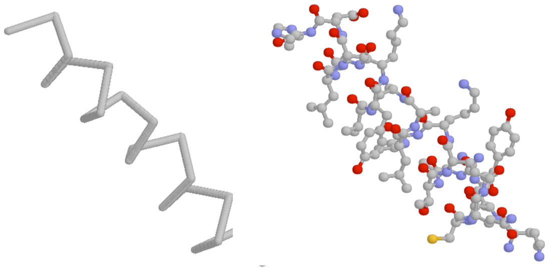
6.4C: The four levels of protein structure

Protein structure can be discussed at four distinct levels. A protein's **primary structure** is two-dimensional - simply the sequence of amino acids in the peptide chain.

Below is a Lewis structure of a short segment of a protein with the sequence CHEM (cysteine - histidine - glutamate - methionine)

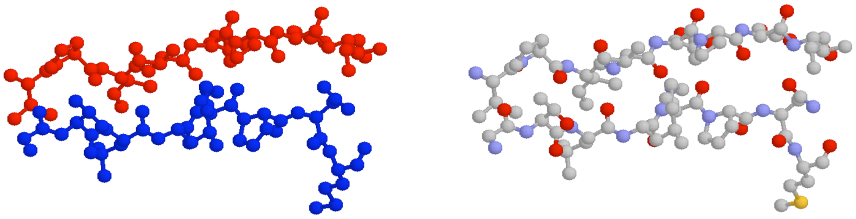


Secondary structure is three-dimensional, but is a local phenomenon, confined to a relatively short stretch of amino acids. For the most part, there are three important elements of secondary structure: helices, β-sheets, and loops. In a helix, the main chain of the protein adopts the shape of a clockwise spiral staircase, and the side chains point out laterally.

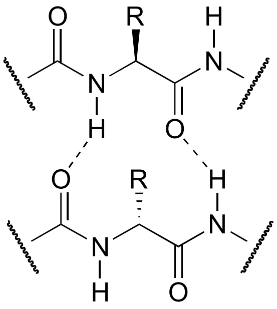


main chain view main chain + side chain

In a β-sheet structure, two sections of protein chain are aligned side-by-side in an extended conformation. The figure below shows two different views of the same β-sheet: in the left-side view, the two regions of protein chain are differentiated by color.

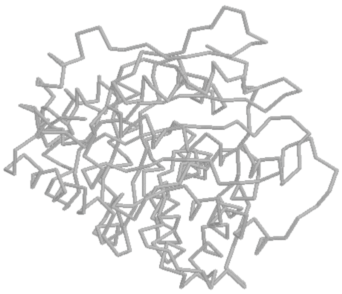


Loops are relatively disordered segments of protein chain, but often assume a very ordered structure when in contact with a second protein or a smaller organic compound. Both the helix and the β-sheet structures are held together by very specific hydrogen-bonding interactions between the amide nitrogen on one amino acid and the carbonyl oxygen on another. The hydrogen bonding pattern in a section of a β-strand is shown below.

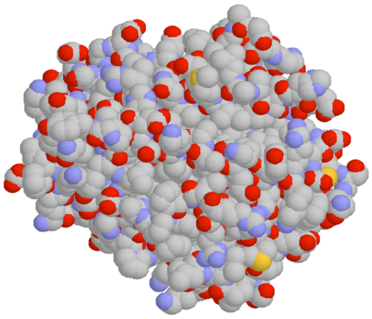


hydrogen bonding in a β-strand

A protein’s **tertiary structure** is the shape in which the entire protein chain folds together in three-dimensional space, and it is this level of structure that provides protein scientists with the most information about a protein’s specific function.

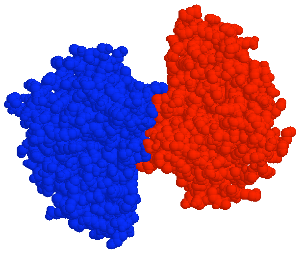


folded protein chain
(main chain view)

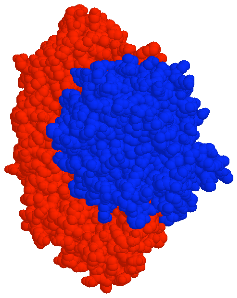


folded protein chain
('space-filling' view)

While a protein's secondary and tertiary structure is defined by how the protein chain folds together, **quaternary structure** is defined by how two or more folded protein chains come together to form a 'superstructure'. Many proteins consist of only one protein chain, or **subunit**, and thus have no quaternary structure. Many other proteins consist of two identical subunits (these are called homodimers) or two non-identical subunits (these are called heterodimers).

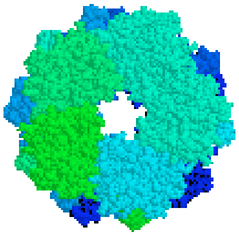


a homodimer

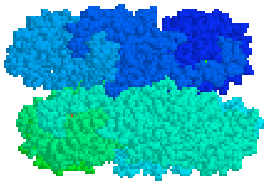


a heterodimer

Quaternary structures can be quite elaborate: below we see a protein whose quaternary structure is defined by ten identical subunits arranged in two five-membered rings, forming what can be visualized as a 'double donut' shape (this is pdb 1l6w).



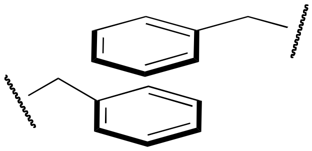
top view



side view

6.4D: The molecular forces that hold proteins together

The question of exactly how a protein 'finds' its specific folded structure out of the vast number of possible folding patterns is still an active area of research. What *is* known, however, is that the forces that cause a protein to fold properly and to remain folded are the same basic noncovalent forces that we talked about in chapter 2: ion-ion, ion-dipole, dipole-dipole, hydrogen bonding, and hydrophobic (van der Waals) interactions. One interesting type of hydrophobic interaction is called 'aromatic stacking', and occurs when two or more planar aromatic rings on the side chains of phenylalanine, tryptophan, or tyrosine stack together like plates, thus maximizing surface area contact.



Hydrogen bonding networks are extensive within proteins, with both side chain and main chain atoms participating. Ionic interactions often play a role in protein structure, especially on the protein surface, as negatively charged residues such as aspartate interact with positively-charged groups on lysine or arginine.

One of the most important ideas to understand regarding tertiary structure is that *a protein, when properly folded, is polar on the surface and nonpolar in the interior*. It is the protein's surface that is in contact with water, and therefore the surface must be hydrophilic in order for the whole structure to be soluble. If you examine a three dimensional protein structure you will see many charged side chains (e.g. lysine, arginine, aspartate, glutamate) and hydrogen-bonding side chains (e.g. serine, threonine, glutamine, asparagine) exposed on the surface, in direct contact with water. Inside the protein, out of contact with the surrounding water, there tend to be many more hydrophobic residues such as alanine, valine, phenylalanine, etc. If a protein chain is caused to come unfolded (through exposure to heat, for example, or extremes of pH), it will usually lose its solubility and form solid precipitates, as the hydrophobic residues from the interior come into contact with water. You can see this phenomenon for yourself if you pour a little bit of vinegar (acetic acid) into milk. The solid clumps that form in the milk are proteins that have come unfolded due to the sudden acidification, and precipitated out of solution.

In recent years, scientists have become increasing interested in the proteins of so-called 'thermophilic' (heat-loving) microorganisms that thrive in hot water environments such as geothermal hot springs. While the proteins in most organisms (including humans) will rapidly unfold and precipitate out of solution when put in hot water, the proteins of thermophilic microbes remain completely stable, sometimes even in water that is just below the boiling point. In fact, these proteins typically only gain full biological activity when in appropriately hot water - at room-temperature they act is if they are 'frozen'. Is the chemical structure of these thermostable proteins somehow unique and exotic? As it turns out, the answer to this question is 'no': the overall three-dimensional structures of thermostable proteins look very much like those of 'normal' proteins. The critical difference seems to be simply that thermostable proteins have more extensive networks of noncovalent interactions, particularly ion-ion interactions on their surface, that provides them with a greater stability to heat. Interestingly, the proteins of 'psychrophilic' (cold-loving) microbes isolated from pockets of water in arctic ice show the opposite characteristic: they have far fewer ion-ion interactions, which gives them greater flexibility in cold temperatures but leads to their rapid unfolding in room temperature water.

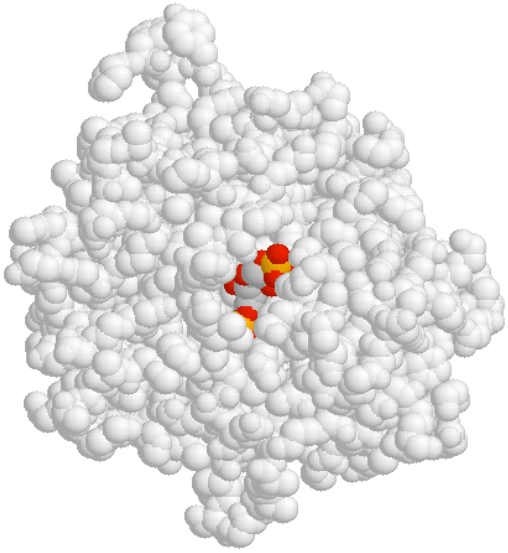
Section 6.5: How enzymes work

So far, what we have learned about protein structure applies to all kinds of proteins with all kinds of functions. It is time now to focus on enzymes, those proteins whose role it is to catalyze the myriad organic (and inorganic) reactions that occur in living things.

We already know, in *very* general terms, what catalysis is: it is the lowering of the activation energy of a reaction that results in rate acceleration. Using slightly different terms to describe the same thing, we can also speak of catalysis as 'transition state stabilization'. Now, we need to think about just *how* an enzyme lowers a reaction's activation energy. We will introduce here some of the most fundamental concepts of enzyme catalysis, and we will continue to encounter specific examples throughout the rest of the text as we learn about different organic reaction mechanisms.

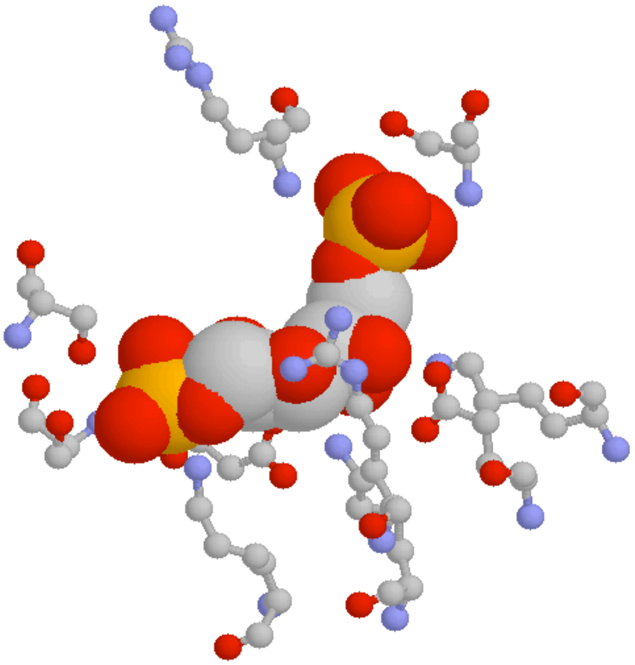
6.5A: The active site

A critical element in the three-dimensional structure of any enzyme is the presence of an '**active site**', which is a pocket, usually located in the interior of the protein, that serves as a docking point for the enzyme's **substrate**(s) ('substrate' is the term that biochemists use for a reactant molecule in an enzyme-catalyzed reaction). It is inside the active site pocket that enzymatic catalysis occurs. Shown below is an image of the glycolytic enzyme fructose-1,6-bisphosphate aldolase, with its substrate bound inside the active site pocket (x-ray crystallographic data are from *Protein Science* **1999**, 8, *291*; pdb 4ALD).

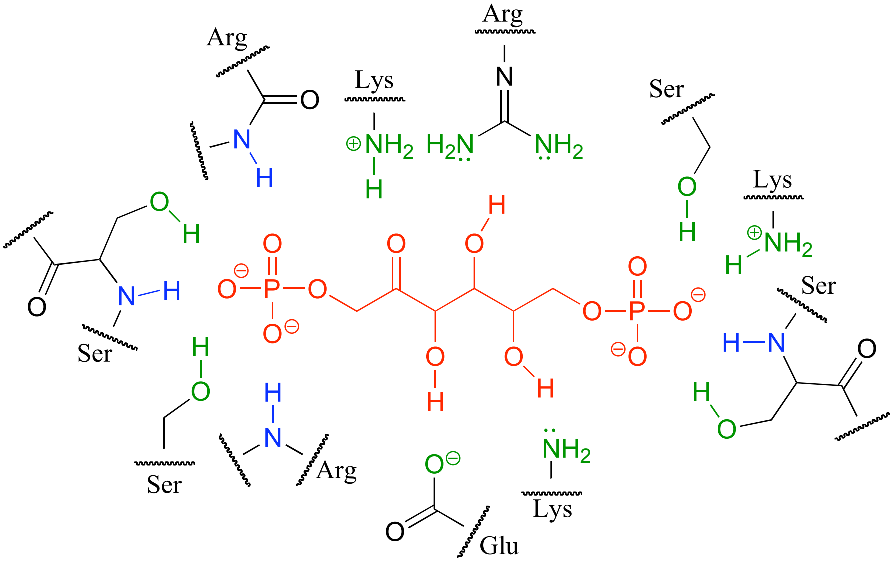


When the substrate binds to the active site, a large number of noncovalent interactions form with the amino acid residues that line the active site. The shape of the active site, and the enzyme-substrate interactions that form as a result of substrate binding, are *specific to the substrate-enzyme pair*: the active site has evolved to 'fit' one particular substrate and to catalyze one particular reaction. Other molecules do not fit in this active site nearly so well as fructose 1,6-bisphosphate.

Here are two close-up views of the same active site pocket, showing some of the specific hydrogen-bonding interactions between the substrate and active site amino acids. The first image below is a three-dimensional rendering directly from the crystal structure data. The substrate is shown in 'space-filling' style, while the active site amino acids are shown in the 'ball and stick' style. Hydrogens are not shown. The color scheme is grey for carbon, red for oxygen, blue for nitrogen, and orange for phosphorus.



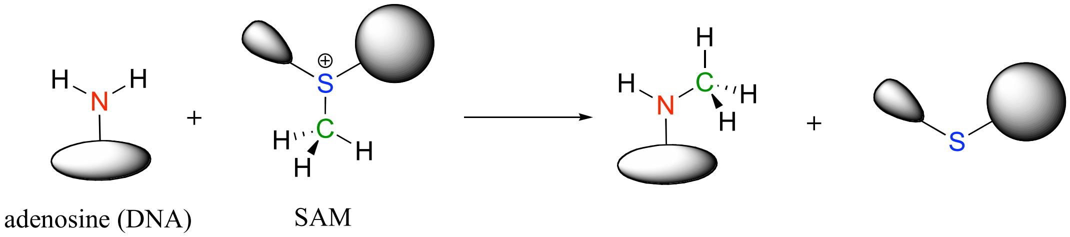
Below is a two-dimensional picture of the substrate (colored red) surrounded by hydrogen-bonding active site amino acids. Notice that both main chain and side chain groups contribute to hydrogen bonding: in this figure, main chain H-bonding groups are colored blue, and side chain H-bonding groups are colored green.



Looking at the last three images should give you some appreciation for the specific manner in which a substrate fits inside its active site.

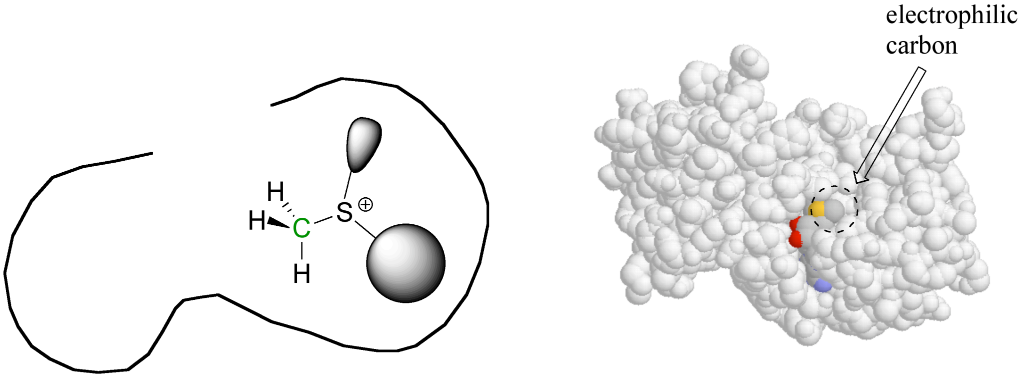
6.5B: Transition state stabilization

One of the most important ways that an enzyme catalyzes any given reaction is through entropy reduction: by bringing order to a disordered situation (remember that entropy is a component of Gibbs Free Energy, and thus a component of the activation energy). Let's turn again to our previous example (from section 6.1C) of a biochemical nucleophilic substitution reaction, the methylation of adenosine in DNA. The reaction is shown below with non-reactive sections of the molecules depicted by variously shaped 'bubbles' for the sake of simplicity.



In order for this reaction to occur, the two substrates (reactants) must come into contact in precisely the right way. If they are both floating around free in solution, the likelihood of this occurring is very small – the entropy of the system is simply too high. In other words, this reaction takes place *very* slowly without the help of a catalyst.

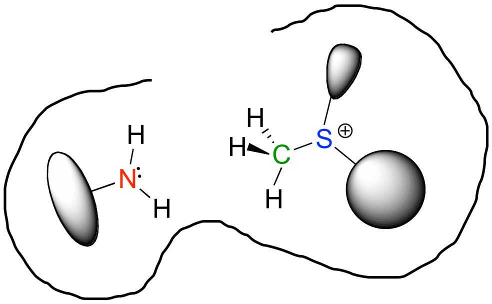
Here's where the enzyme's active site pocket comes into play. It is lined with various functional groups from the amino acid main and side chains, and has a very specific three-dimensional architecture that has evolved to bind to both of the substrates. If the SAM molecule, for example, diffuses into the active site, it can replace its (favorable) interactions with the surrounding water molecules with (even more favorable) new interactions with the functional groups lining the active site.



SAM bound in active site

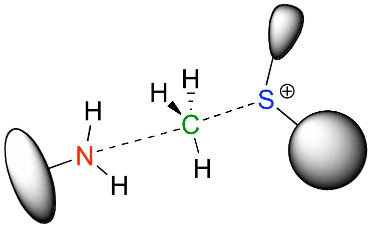
In a sense, SAM is moving from one solvent (water) to another 'solvent' (the active site), where many new energetically favorable interactions are possible. Remember: these new contacts between SAM and the active site groups are *highly specific* to SAM and SAM alone – no other molecule can 'fit' so well in this precise active site environment, and thus no other molecule will be likely to give up its contacts to water and bind to the active site.

The second substrate also has a specific spot reserved in the active site. (Because in this case the second substrate is a small segment of a long DNA molecule, the DNA-binding region of the active site is more of a 'groove' than a 'pocket').



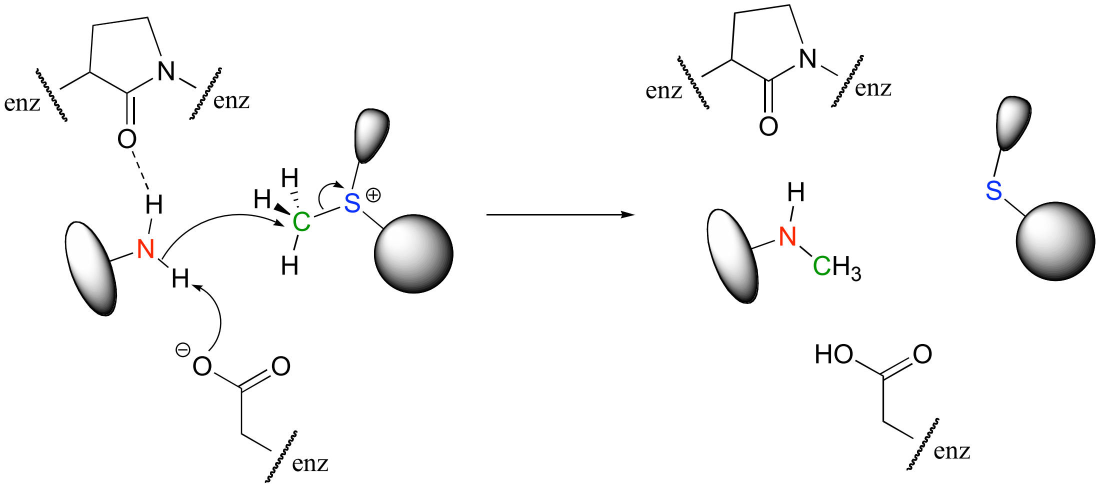
So now we have both substrates bound in the active site. But they are not just bound in any random orientation – they are specifically positioned relative to one another so that the nucleophilic nitrogen is held very close to the electrophilic carbon, with a free path of attack. What used to be a very disordered situation – two reactants diffusing freely in solution – is now a very highly ordered situation, with everything set up for the reaction to proceed. This is what is meant by entropy reduction: the entropic component of the energy barrier has been lowered.

Looking a bit deeper, though, it is not really the noncovalent interaction between enzyme and *substrate* that are responsible for catalysis. Remember: all catalysts, enzymes included, accelerate reactions by lowering the energy of the *transition state*. With this in mind, it should make sense that the primary job of an enzyme is to maximize favorable interactions with the transition state, *not* with the starting substrates. This does not imply that enzyme-substrate interactions are not strong, rather that enzyme-TS interactions are far *stronger*, often by several orders of magnitude. Think about it this way: if an enzyme were to bind to (and stabilize) its substrate(s) more tightly than it bound to (and stabilized) the transition state, it would actually *slow down* the reaction, because it would be *increasing* the energy difference between starting state and transition state. ***The enzyme has evolved to maximize favorable noncovalent interactions to the transition state***: in our example, this is the state in which the nucleophilic nitrogen is already beginning to attack the electrophilic carbon, and the carbon-sulfur bond has already begun to break.

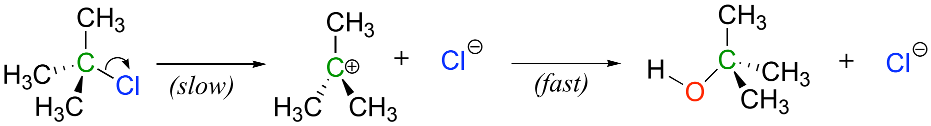


enzyme binds best to the
transition state

In many enzymatic reactions, certain active site amino acid residues contribute to catalysis by *increasing the reactivity of the substrates*. Often, the catalytic role is that of acid and/or base. In our DNA methylation example, the nucleophilic nitrogen is deprotonated by a nearby aspartate side chain as it begins its nucleophilic attack on the methyl group of SAM. We will study nucleophilicity in greater detail in chapter 8, but it should make intuitive sense that deprotonating the amine increases the electron density of the nitrogen, making it *more nucleophilic*. Notice also in the figure below that the main chain carbonyl of an active site proline forms a hydrogen bond with the amine, which also has the effect of increasing the nitrogen's electron density and thus its nucleophilicity (*Nucleic Acids Res*. **2000**, *28*, 3950).

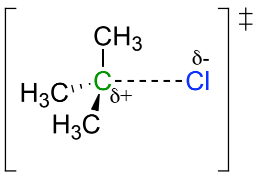


How does our picture of enzyme catalysis apply to multi-step reaction mechanisms? Although the two-step nucleophilic substitution reaction between *tert*-butyl chloride and hydroxide (section 6.1C) is not a biologically relevant process, let’s pretend just for the sake of illustration that there is a hypothetical enzyme that catalyzes this reaction.



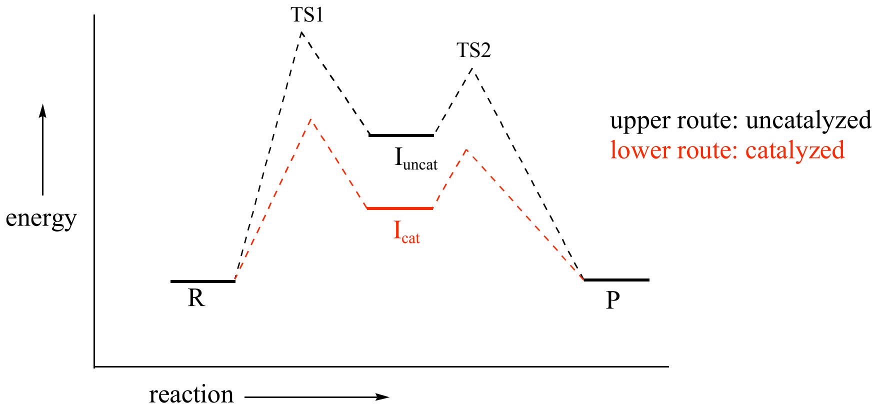
The same basic principles apply here: the enzyme binds best to the transition state. But therein lies the problem: there are two transition states! To which TS does the enzyme maximize its contacts?

Recall that the first step – the loss of the chloride leaving group to form the carbocation intermediate – is the slower, rate-limiting step. It is this step that our hypothetical enzyme needs to accelerate if it wants to accelerate the overall reaction, and it is thus the energy of TS1 that needs to be lowered.



enzyme maximizes interactions with TS1

By Hammond's postulate, we also know that the intermediate I is a close approximation of TS1. So the enzyme, by stabilizing the intermediate, will also stabilize TS1 (as well as TS2) and thereby accelerate the reaction.

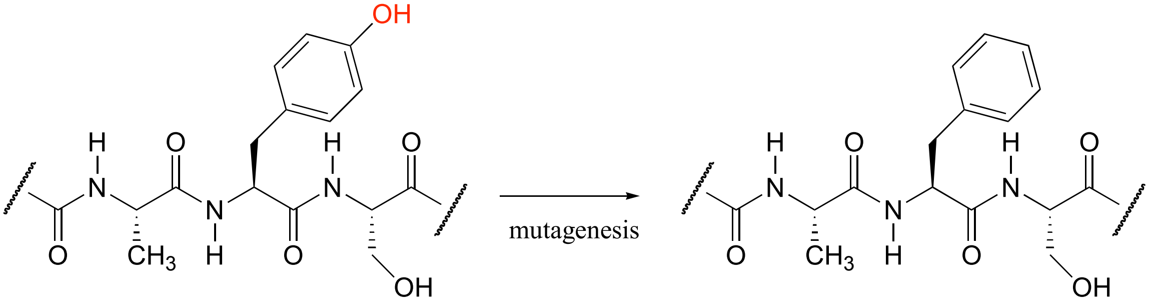


If you read scientific papers about enzyme mechanisms, you will often see researchers discussing how an enzyme stabilizes a reaction intermediate. By virtue of Hammond's postulate, they are, at the same time, talking about how the enzyme lowers the energy of the transition state.

An additional note: although we have in this section been referring to SAM as a 'substrate' of the DNA methylation reaction, it is also often referred to as a **coenzyme**, or **cofactor**. These terms are used to describe small (relative to protein and DNA) biological organic molecules that bind specifically in the active site of an enzyme and help the enzyme to do its job. In the case of SAM, the job is methyl group donation. In addition to SAM, we will see many other examples of coenzymes in the coming chapters, a number of which - like ATP (adenosine triphosphate), coenzyme A, thiamine, and flavin - you have probably heard of before. The full structures of some common coenzymes are shown in table 6 in the tables section.

6.5C: Site-directed mutagenesis

Biochemists frequently use a powerful technique called **site-directed mutagenesis** in order to learn more about the role played by a particular amino acid residue or group of residues in a protein of interest. The technical details of site-directed mutagenesis are beyond the scope of this text, but the end result is not hard to understand: essentially, a researcher is able to change any amino acid in a protein to any other amino acid. If we are interested in the role played by the phenol group in an active site tyrosine, for example, we might use site-directed mutagenesis to produce a mutant protein in which this tyrosine has been changed to a phenylalanine, thus preserving the phenyl group but specifically deleting the phenol.

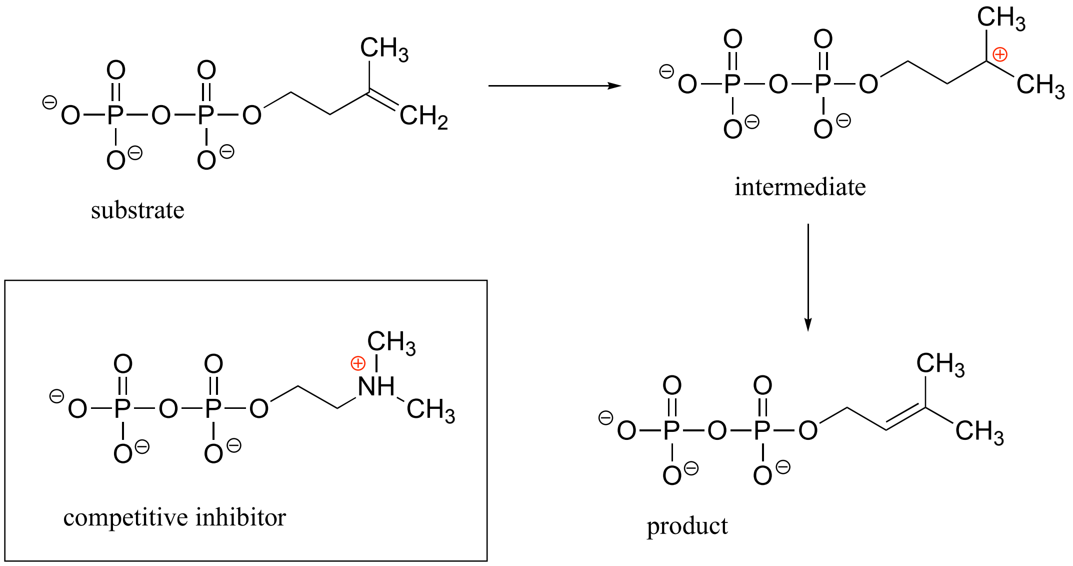


If the mutant protein behaves differently from the 'wild type' (natural) protein - if it loses all activity, for example - then we can infer that the phenol group plays some important role. In some cases, researchers have even been able to change the product or substrate specificity of an enzymatic reaction using site-directed mutagenesis. The technique has become an indispensable tool for researchers studying enzyme-catalyzed organic reactions.

6.5D: Enzyme inhibitors

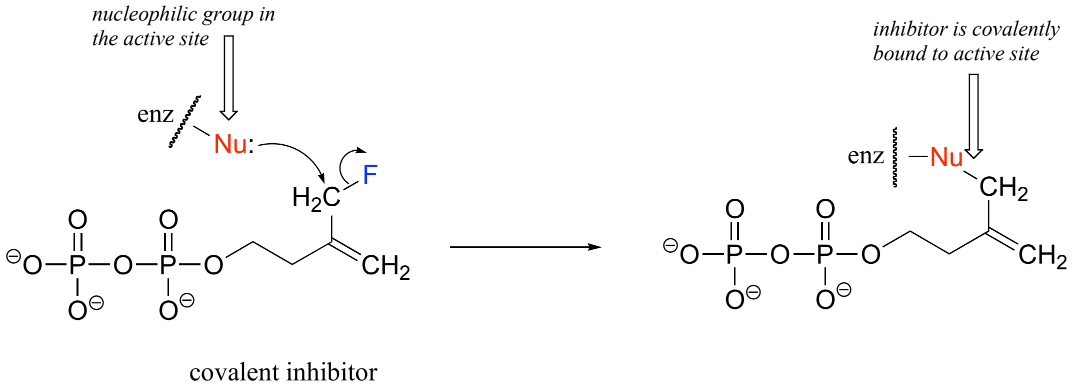
The action of enzymes can be slowed or shut down completely by agents called **inhibitors**. Although there are several different modes of enzyme inhibition, we will focus here on two of the most important: **competitive inhibition** and **irreversible inhibition**. In competitive inhibition, a molecule with a structure that closely resembles that of the substrate, intermediate, or product of a particular enzymatic reaction is able to enter and bind (through noncovalent interactions) to the enzyme's active site. This 'imposter' molecule, as long as it stays in the active site, will prevent the real substrate from entering and therefore will prevent the enzyme from catalyzing its reaction. The potency of a competitive inhibitor depends on how well it binds to the active site.

Shown below is an example of a competitive inhibitor for an enzyme called isopentenyl diphosphate isomerase (we'll study the reaction it catalyzes in section 15.3A). Looking at the structure of the inhibitor, you can see its resemblance to both the substrate and the product of the reaction - but you can also see that, with its positive charge, *it more closely resembles the cationic reaction intermediate*.



This makes perfect sense - remember that enzymes have evolved to bind to the transition state of their reactions more tightly than to either the substrate or the product, and by Hammond's postulate we know that the intermediate resembles the transition state more than the substrate or product do. Therefore, an organic chemist who wants to design a potent competitive inhibitor would be wise to try to imitate the structure of the reaction intermediate.

A irreversible inhibitor (also known as a **suicide substrate**) is also designed to be recognized by its target enzyme, but there is one critical difference. A competitive inhibitor is free to diffuse in and out of the active site (just like an actual enzymatic substrate or product). A covalent inhibitor, in contrast, will form a strong *covalent* bond to an active site residue after entering the site: once it is in, it will never come out, and thus it has *irreversible* inactivated the enzyme. Shown below is a covalent inhibitor for isopentenyl diphosphate isomerase:

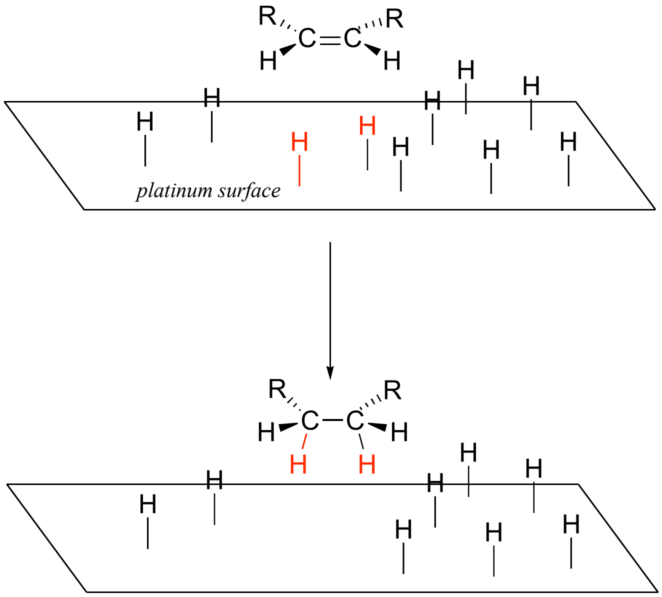

nucleophilic group in the active site
enz
Nu:
F
inhibitor is covalently bound to active site
enz
Nu
covalent inhibitor

As you can see, the inhibitor resembles the natural substrate, but has a fluorine leaving group on its methyl carbon. The methyl carbon is now an electrophilic 'bait' for any nucleophilic group that happens to be nearby in the active site, and the result is irreversible, covalent inactivation of the enzyme - the inhibitor is locked inside the active sight for good. (Both of these inhibitor examples were first described in *Biochemistry* **1988**, 27, 7315.)

Many drugs work by inhibiting enzymes. The antibiotic penicillin, for example, is an irreversible inhibitor of a bacterial enzyme involved in the construction of cell walls (we will study the action of penicillin in more detail in section 12.5C). The 'statin' family of cholesterol-lowering drugs are all competitive inhibitors of an enzyme called HMG-CoA reductase (section 16.4D), which catalyzes a chemical step in the synthesis of cholesterol.

6.5E: Catalysts in the laboratory

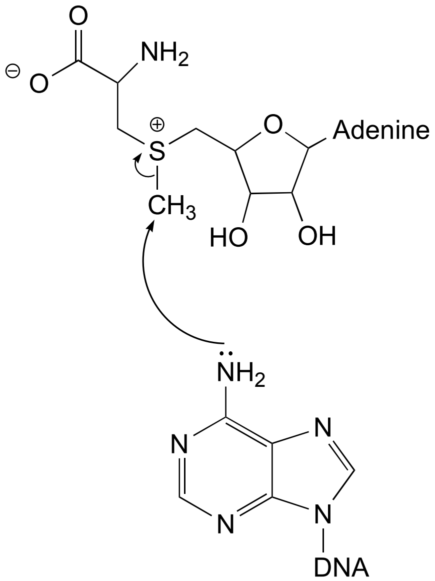
Although we are focusing here on biological catalysis, the role of catalysis in chemistry is by no means limited to enzymes. Chemists synthesizing organic molecules in the laboratory have a wide and ever-growing arsenal of non-biological catalytic strategies at their disposal, a few of which we will see in various places in this text. Often, catalysis in the lab is as simple as adding acid or base to the reaction mixture, in situations where changing the protonation state of a functional group leads to a faster reaction. Metals are frequently used in catalytic roles: one such example that we will discuss later (section 16.5D) is the addition of hydrogen to the double bond of an alkene, a reaction that takes place on the catalytic surface of a metal such as platinum or palladium.

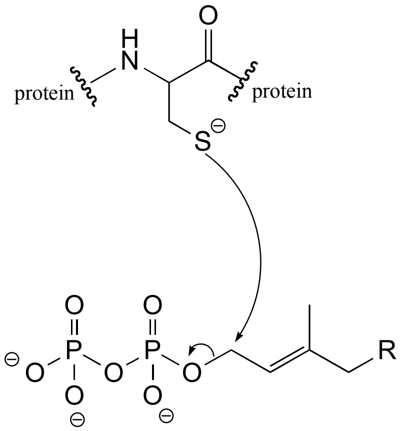


When it comes to efficiency and specificity of catalysis, enzymes for the most part still far outperform their human-designed laboratory counterparts, but most enzymes are inherently fragile and not very effective outside of an aqueous, pH-neutral, 37° environment. In addition, an enzyme catalyzes the specific reaction that it has evolved to catalyze, which often is different from the reaction a human chemist wants to carry out. For these reasons, the design and synthesis of large catalytic molecules that mimic the architecture of enzyme active sites has become a very active research field in recent years. Researchers are also exploring ways to use robust, heat-stable enzymes from 'extremophilic' microbes (such as those that live in the boiling water of hot springs) as laboratory catalysts, and are also trying to selectively change the substrate or product specificity of enzymes through the use of site-directed mutagenesis and other genetic engineering techniques.

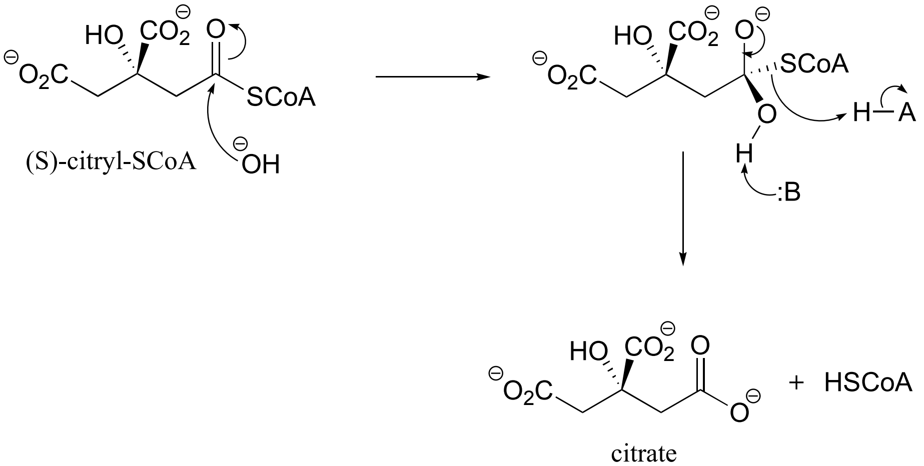
Chapter 6 problems

P6.1: In the reactions illustrated below, identify the functional groups which are acting as nucleophile, electrophile, and leaving group.





P6.2: Below is a reaction in the citric acid (Krebs) cycle.

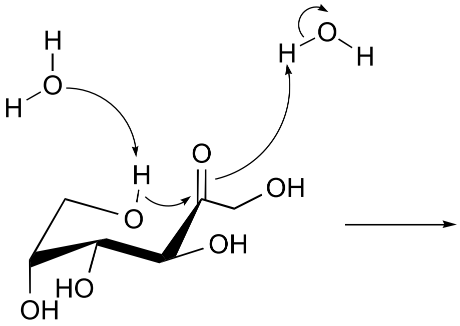


a) Draw pictures illustrating the bond-forming and bond-breaking events at each transition state.

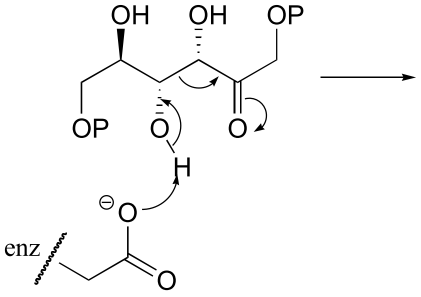
b) Identify the nucleophile and the electrophile in the first step.

P6.3: Illustrated below are individual mechanistic steps in some important biochemical reactions that we will be studying later. For each step, draw the transition state, and also the products or intermediate species that would form according to the electron-movement arrows given. Be sure to include all formal charges.

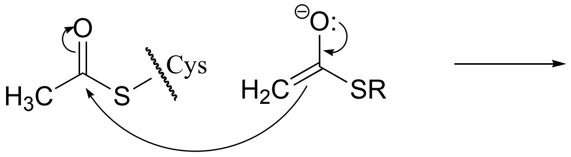
a)



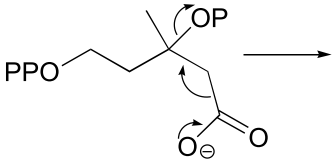
b)



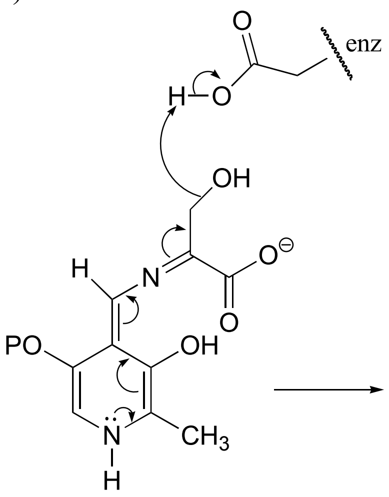
c)



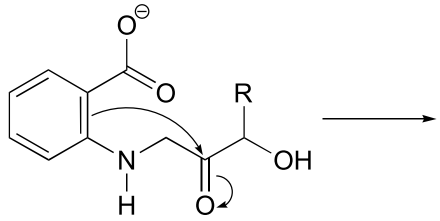
d)



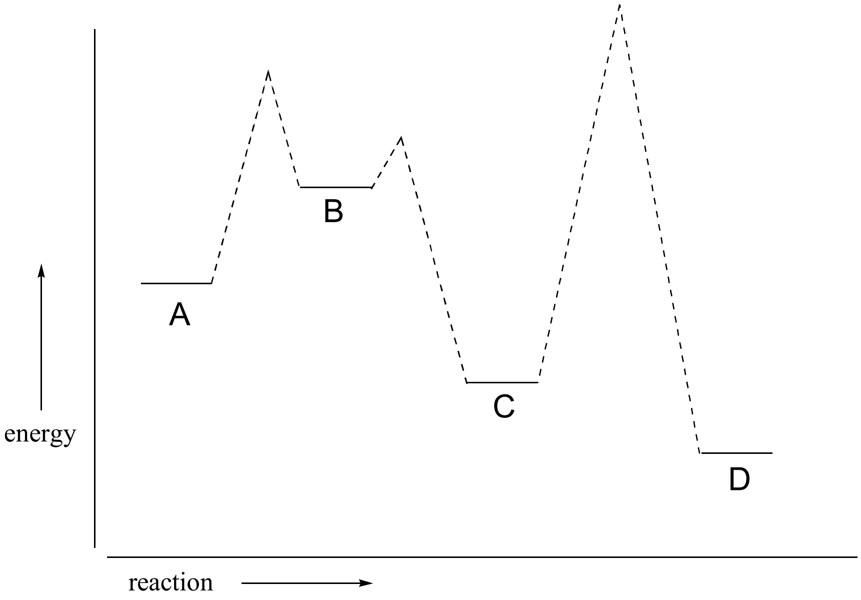
e)



f)



P6.4: Below is an energy diagram for a hypothetical reaction. What is the rate limiting step?



P6.5: Draw a stereochemically accurate Lewis structure of the peptide with the amino acid sequence CKTV.

Chapter 7

Acid-base reactions

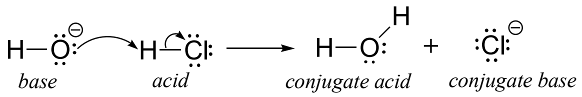
Introduction

In the previous chapter, we introduced many of the fundamental concepts that will form the foundation of your understanding of how organic molecules react. Now, we begin our study of organic reactivity in earnest, beginning with the simplest reaction type: the transfer of a proton from an acid to a base. After reviewing some basic ideas about acid-base equilibria with which you are probably already familiar from General Chemistry, we will dive into some very challenging new waters, as we attempt to use our understanding of organic structure to predict how different organic functional groups are likely to react in an acid-base context. Many of the ideas that are introduced in this chapter, though perhaps difficult to grasp as first, will be crucial to understanding not only acid base chemistry but all of the other organic reaction types that we will see throughout the remainder of the text.

Section 7.1: The 'basic' idea of an acid-base reaction

7.1A: The Brønsted-Lowry definition of acidity

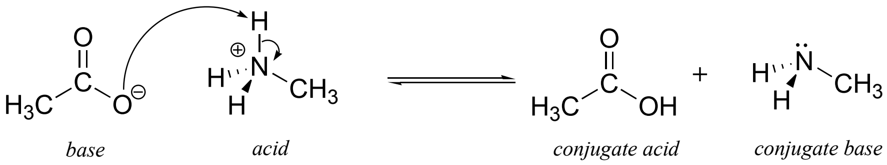
We'll begin our discussion of acid-base chemistry with a couple of essential definitions. The first of these definitions was proposed in 1923 by the Danish chemist Johannes Brønsted and the English chemist Thomas Lowry, and has come to be known as the Brønsted-Lowry definition of acids and bases. An acid, by the Brønsted-Lowry definition, is a species which is able to donate a proton (H^+), while a base is a proton acceptor. We have already discussed in the previous chapter one of the most familiar examples of a Brønsted-Lowry acid-base reaction, between hydrochloric acid and hydroxide ion:



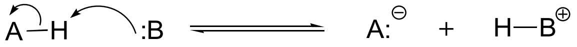
In this reaction, a proton is transferred from HCl (the acid, or proton *donor*) to hydroxide (the base, or proton *acceptor*). As we learned in the previous chapter, curved arrows depict the movement of electrons in this bond-breaking and bond-forming process.

After a Brønsted-Lowry acid donates a proton, what remains – in this case, a chloride ion – is called the **conjugate base**. Chloride is thus the conjugate base of hydrochloric acid. Conversely, when a Brønsted-Lowry base accepts a proton it is converted into its **conjugate acid** form: water is thus the conjugate acid of hydroxide.

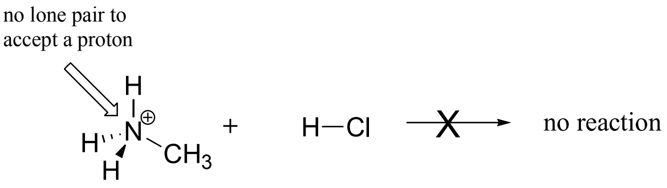
We can also talk about conjugate acid/base pairs: the two acid/base pairs involved in our first reaction are hydrochloric acid/chloride and hydroxide/water. In this next acid-base reaction, the two pairs involved are acetate/acetic acid and methyl ammonium/methylamine:



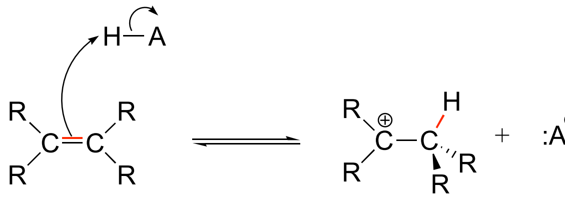
Throughout this text, we will often use the abbreviations HA and :B in order to refer in a general way to acidic and basic reactants:



In order to act as a proton acceptor, a base must have a reactive pair of electrons. In all of the examples we shall see in this chapter, this pair of electrons is a non-bonding lone pair, usually (but not always) on an oxygen, nitrogen, sulfur, or halogen atom. When acetate acts as a base in the reaction shown above, for example, one of its oxygen lone pairs is used to form a new bond to a proton. The same can be said for an amine acting as a base. Clearly, methyl ammonium ion cannot act as a base – it does not have a reactive pair of electrons with which to accept a new bond to a proton.



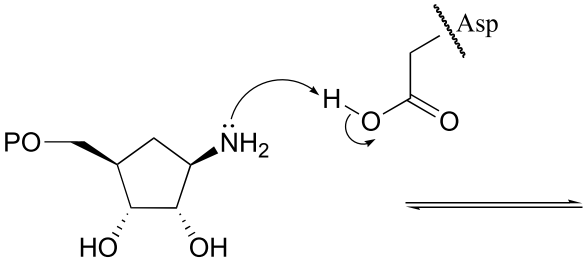
Later, in chapter 15, we will see several examples where the (relatively) reactive pair of electrons in a π bond act in a basic fashion.



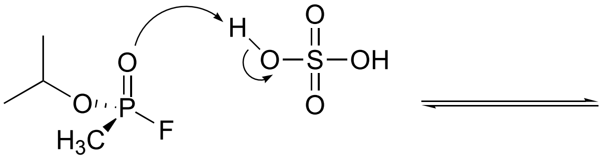
In this chapter, we will concentrate on those bases with non-bonding (lone pair) electrons.

Exercise 7.1: Draw structures for the missing conjugate acids or conjugate bases in the reactions below.

a)

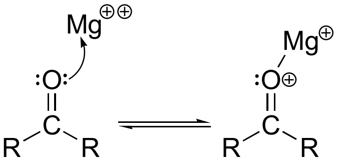


b)



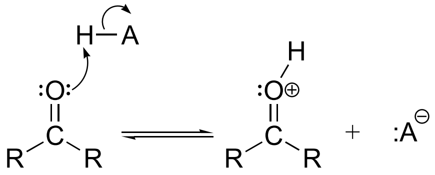
7.1B: The Lewis definition of acidity

The Brønsted-Lowry picture of acids and bases as proton donors and acceptors is not the only definition in common use. A broader definition is provided by the **Lewis theory** of acids and bases, in which a Lewis acid is an electron-pair acceptor and a Lewis base is an electron-pair donor. This definition covers Brønsted-Lowry proton transfer reactions, but also includes reactions in which no proton transfer is involved. The interaction between a magnesium cation (Mg^{+2}) and a carbonyl oxygen is a common example of a Lewis acid-base reaction. The carbonyl oxygen (the Lewis base) donates a pair of electrons to the magnesium cation (the Lewis acid).



As we will see in chapter 11 when we begin the study of reactions involving carbonyl groups, this interaction has the very important effect of increasing the polarity of the carbon-oxygen double bond.

The Brønsted-Lowry equivalent of the reaction above is simply protonation of the carbonyl group. This, too, has the effect of increasing the polarity of the carbonyl double bond.

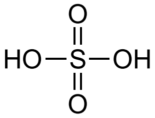


While it is important to be familiar with the Lewis definition, the focus throughout the remainder of this chapter will be on acid-base reactions of the Brønsted-Lowry type, where an actual proton transfer event takes place.

Section 7.2: Comparing the acidity and basicity of organic functional groups– the acidity constant

7.2A: Defining K_a and pK_a

You are no doubt aware that some acids are stronger than others. Sulfuric acid is strong enough to be used as a drain cleaner, as it will rapidly dissolve clogs of hair and other organic material.

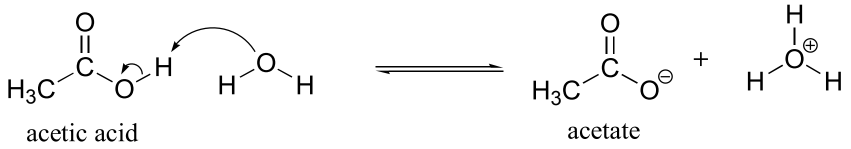


sulfuric acid

Not surprisingly, concentrated sulfuric acid will also cause painful burns if it touches your skin, and permanent damage if it gets in your eyes (there's a good reason for those safety goggles you wear in chemistry lab!). Acetic acid (vinegar), will also burn your skin and eyes, but is not nearly strong enough to make an effective drain cleaner. Water, which we know can act as a proton donor, is obviously not a very strong acid. Even hydroxide ion could *theoretically* act as an acid – it has, after all, a proton to donate – but this is not a reaction that we would normally consider to be relevant in anything but the most extreme conditions.

The relative acidity of different compounds or functional groups – in other words, their relative capacity to donate a proton to a common base under identical conditions – is quantified by a number called the **dissociation constant**, abbreviated $\mathbf{K_a}$. The common base chosen for comparison is water.

We will consider acetic acid as our first example. When a small amount of acetic acid is added to water, a proton-transfer event (acid-base reaction) occurs to some extent.



Notice the phrase 'to some extent' – this reaction does *not* run to completion, with all of the acetic acid converted to acetate, its conjugate base. Rather, a *dynamic equilibrium* is reached, with proton transfer going in both directions (thus the two-way arrows) and finite concentrations of all four species in play. The nature of this equilibrium situation, as you recall from General Chemistry, is expressed by an equilibrium constant, K_{eq}.

$$K_{eq} = \frac{[\text{products}]}{[\text{reactants}]} = \frac{[CH_3COO^-]\,[H_3O^+]}{[CH_3COOH]\,[H_2O]}$$

We added a small amount of acetic acid to a large amount of water: water is the *solvent* for this reaction. Therefore, in the course of the reaction, the concentration of water (approximately 55.6 mol/L) changes very little, and can be treated as constant. The acid dissociation constant, or K_a, for acetic acid is defined as:

$$K_a = K_{eq}[H_2O] = \frac{[CH_3COO^-]\,[H_3O^+]}{[CH_3COOH]} = 1.75 \times 10^{-5}$$

In more general terms, the dissociation constant for a given acid is expressed as:

$$K_a = \frac{[A^-]\,[H_3O^+]}{[HA]} \quad \textit{or} \quad K_a = \frac{[A]\,[H_3O^+]}{[HA^+]}$$

The first expression applies to a neutral acid such as like HCl or acetic acid, while the second applies to a cationic acid like ammonium (NH_4^+).

The value of $K_a = 1.75 \times 10^{-5}$ for acetic acid is very small - this means that very little dissociation actually takes place, and there is much more acetic acid in solution at equilibrium than there is acetate ion. Acetic acid is a relatively weak acid, at least when compared to sulfuric acid ($K_a = 10^9$) or hydrochloric acid ($K_a = 10^7$), both of which undergo essentially complete dissociation in water.

A number like 1.75×10^{-5} is not very easy either to say or to remember. Chemists have therefore come up with a more convenient term to express relative acidity: the **pK_a** value.

$pK_a = -\log K_a$

Doing the math, we find that the pK_a of acetic acid is 4.8. The use of pK_a values allows us to express the acidity of common compounds and functional groups on a numerical scale of about –10 (very strong acid) to 50 (not acidic at all). Table 7 at the end of the text lists exact or approximate pK_a values for different types of protons that you are likely to encounter in your study of organic and biological chemistry. Looking at Table 7, you see that the pK_a of carboxylic acids are in the 4-5 range, the pK_a of sulfuric acid is –10, and the pK_a of water is 15.7. Alkenes and alkanes, which are not acidic at all, have pK_a values above 30. *The lower the pK_a value, the stronger the acid.*

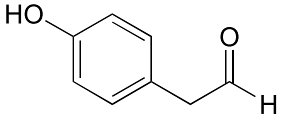
It is important to realize that pK_a is not at all the same thing as pH: the former is an inherent property of a compound or functional group, while the latter is the measure of the hydronium ion concentration in a particular aqueous solution:

$pH = -\log [H_3O^+]$

Any particular acid will always have the same pK_a (assuming that we are talking about an aqueous solution at room temperature) but different aqueous solutions of the acid could have different pH values, depending on how much acid is added to how much water.

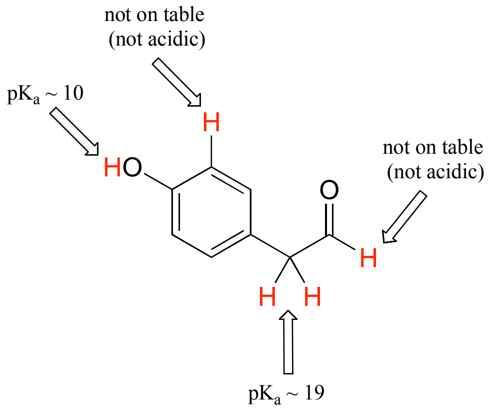
Our table of pK_a values will also allow us to compare the strengths of different bases by comparing the pK_a values of their conjugate acids. The key idea to remember is this: *the stronger the conjugate acid, the weaker the conjugate base*. Sulfuric acid is the strongest acid on our list with a pK_a value of –10, so HSO_4^- is the weakest conjugate base. You can see that hydroxide ion is a stronger base than ammonia (NH_3), because ammonium (NH_4^+, $pK_a = 9.2$) is a stronger acid than water ($pK_a = 15.7$).

While Table 7 provides the pK_a values of only a limited number of compounds, it can be very useful as a starting point for estimating the acidity or basicity of just about any organic molecule. Here is where your familiarity with organic functional groups will come in very handy. What, for example, is the pK_a of cyclohexanol? It is not on the table, but as it is an alcohol it is probably somewhere near that of ethanol ($pK_a = 16$). Likewise, we can use Table 7 to predict that para-hydroxyphenyl acetaldehyde, an intermediate compound in the biosynthesis of morphine, has a pK_a in the neighborhood of 10, close to that of our reference compound, phenol.



p=hydroxyphenyl acetaldehyde

Notice in this example that we need to evaluate the potential acidity at *four* different locations on the molecule.



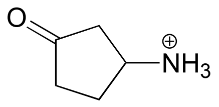
Aldehyde and aromatic protons are not at all acidic (pK_a values are above 40 – not on our table). The two protons on the carbon next to the carbonyl are slightly acidic, with pK_a values around 19-20 according to the table. The most acidic proton is on the phenol group, so if the compound were to be subjected to a single molar equivalent of strong base, this is the proton that would be donated.

As you continue your study of organic chemistry, it will be a very good idea to commit to memory the approximate pK_a ranges of some important functional groups, including water, alcohols, phenols, ammonium, thiols, phosphates, carboxylic acids and carbons next to carbonyl groups (so-called α-carbons). These are the groups that you are most likely to see acting as acids or bases in biological organic reactions.

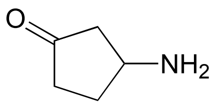
A word of caution: when using the pK_a table, be absolutely sure that you are considering the correct conjugate acid/base pair. If you are asked to say something about the basicity of ammonia (NH_3) compared to that of ethoxide ion ($CH_3CH_2O^-$), for example, the relevant pK_a values to consider are 9.2 (the pK_a of ammonium ion) and 16 (the pK_a of ethanol). From these numbers, you know that ethoxide is the stronger base. Do not make the mistake of using the pK_a value of 38: this is the pK_a of ammonia *acting as an acid*, and tells you how basic the NH_2^- ion is (very basic!)

Exercise 7.2: Using the pK_a table, estimate pK_a values for the most acidic group on the compounds below, and draw the structure of the conjugate base that results when this group donates a proton.

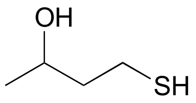
a)



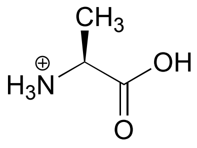
b)



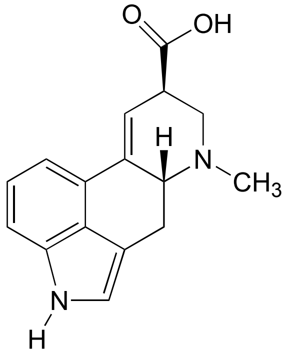
c)



d)



e)



Lysergic acid

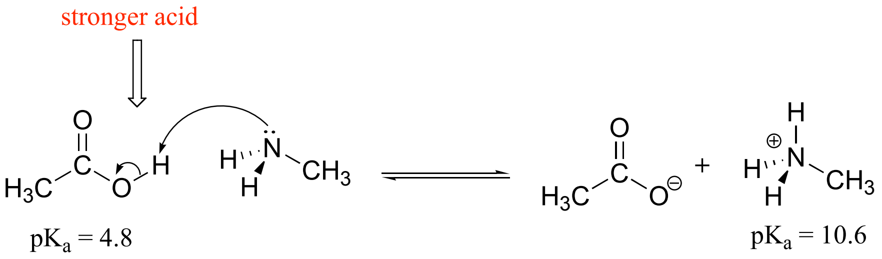
7.2B: Using pK_a values to predict reaction equilibria

By definition, the pK_a value tells us the extent to which an acid will react with water as the base, but by extension, we can also calculate the equilibrium constant for a reaction between any acid-base pair. Mathematically, it can be shown that:

K_{eq} (for the acid base reaction in question) = $10^{\Delta pKa}$

where $\Delta pK_a = pK_a$ of product acid minus pK_a of reactant acid

Consider a reaction between methylamine and acetic acid:



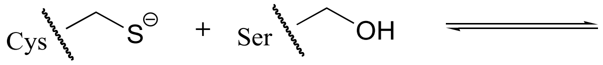
First, we need to identify the acid species on either side of the equation. On the left side, the acid is of course acetic acid, while on the right side the acid is methyl ammonium. The specific pK_a values for these acids are not on our very generalized pK_a table, but are given in the figure above. Without performing any calculations, you should be able to see that this equilibrium lies far to the right-hand side: acetic acid has a lower pK_a, is a stronger acid, and thus it wants to give up its proton more than methyl ammonium does. Doing the math, we see that

$$K_{eq} = 10^{\Delta pKa} = 10^{(10.6 - 4.8)} = 10^{5.8} = 6.3 \times 10^{5}$$

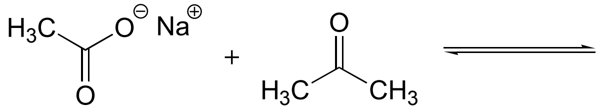
So K_{eq} is a very large number (much greater than 1) and the equilibrium lies far to the right-hand side of the equation, just as we had predicted. If you had just wanted to approximate an answer without bothering to look for a calculator, you could have noted that the difference in pK_a values is approximately 6, so the equilibrium constant should be somewhere in the order of 10^6, or one million. Using the pK_a table in this way, and making functional group-based pK_a approximations for molecules for which we don't have exact values, we can easily estimate the extent to which a given acid-base reaction will proceed.

Exercise 7.3 Show the products of the following acid-base reactions, and estimate the value of K_{eq}.

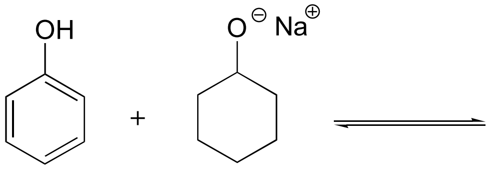
a)



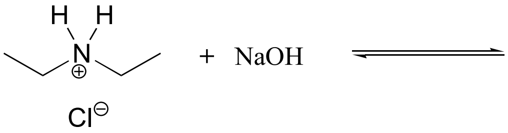
b)



c)



d)



7.2C: pK_a and pH: the Henderson-Hasselbalch equation

It was mentioned previously that pK_a and pH are not the same thing - however, they *are* related. Probably the most commonly-used equation relating the two concepts is known as the Henderson-Hasselbalch equation (which is really just a rearrangement of the equation that defines pK_a):

$$pH = pK_a + \log\left(\frac{[A^-]}{[HA]}\right) \quad \textit{or} \quad pH = pK_a + \log\left(\frac{[A]}{[HA^+]}\right)$$

If an acid is exactly 50% dissociated in aqueous solution, then the concentration of A^- is equal to the concentration of HA, and the fraction term in the Henderson-Hasselbalch equation is equal to 1. Because log(1) = 0, it follows that pH = pK_a at these conditions. This is very useful: it means that *the pK_a of an acid is equal to the pH at which it is 50% disassociated* (deprotonated). For example, if a 0.1 M aqueous solution of acetic acid is brought to pH 4.8 by the addition of sodium hydroxide, we know by the Henderson-Hasselbalch equation that the solution is now 0.05 M in acetic acid and 0.05 M in acetate ion.

The above situation – an aqueous solution that contains equal or nearly equal concentrations of a weak acid and its conjugate base – should sound very familiar to you. This is a buffer! As we discuss the organic reactions that occur in living things, it will be very important to always keep in mind that most of these reactions are occurring in an aqueous solution that is buffered to approximately pH 7.3 (the exceptions are reactions occuring in lysosomes and endosomes, specialized organelles in eukaryotic cells that maintain a slightly acidic interior). The buffer in a living cell is not composed of acetic acid/acetate mixture, or course – that would make the pH far too acidic. Rather, the chemistry of life occurs in a buffer that consists of a mixture of various phosphate and ammonium compounds.

So, what does the side chain of an aspartate amino acid residue look like if it is on the surface of a protein in an aqueous solution buffered to pH 7.0? Is it protonated or deprotonated? With an approximate pK_a of 3.9, the Henderson-Hasselbalch equation tells us that the side chain should be greater than 99% deprotonated:

$7.0 = 3.9 + \log([A-] / [HA])$
$([A-] / [HA]) = 1.3 \times 10^3$

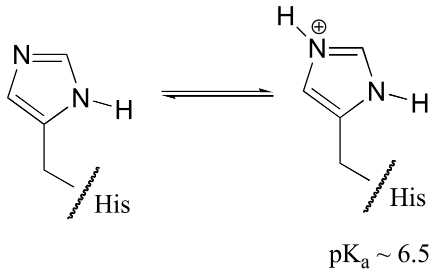
$\Rightarrow [A^-] >> [HA]$

What about the amino group on a lysine side chain? With an approximate pK_a of 10.8, it should be >99% protonated, in the positively-charged, ammonium form:

$7.0 = 10.8 + \log([A] / [HA^+])$
$([A] / [HA^+]) = 1.6 \times 10^{-4}$

$\Rightarrow [HA^+] >> [A]$

So, in an aqueous solution buffered at pH 7, carboxylic acid groups can be expected to be essentially 100% deprotonated and negatively charged (ie. in the carboxylate form), and amine groups essentially 100% protonated and positively charged (ie. in the ammonium form). Alcohols are fully protonated and neutral at pH 7, as are thiols. The imidizole group on the histidine side chain has a pK_a near 7, and thus exists in physiological solutions as mixture of both protonated and deprotonated forms.



Exercise 7.4: If an arginine side chain in a protein is exposed to buffer with pH = 8.5, to what extent (expressed as a percentage) is it deprotonated?

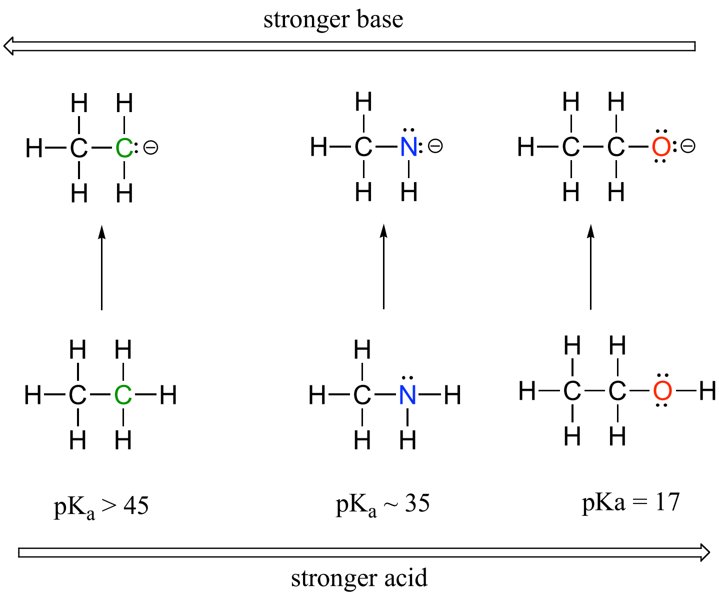
Exercise 7.5: A small amount of acetic acid is dissolved in a buffer of pH = 5.2. What percent of the acetic acid molecules are charged?

Section 7.3: Structural effects on acidity and basicity

Now that we know how to quantify the strength of an acid or base, our next job is to gain an understanding of the fundamental reasons behind *why* one compound is more acidic or more basic than another. This is a big step: we are, for the first time, taking our knowledge or organic *structure* and applying it to a question of organic *reactivity*. Many of the ideas that we'll see for the first here will continue to apply throughout the text as we tackle many other organic reaction types.

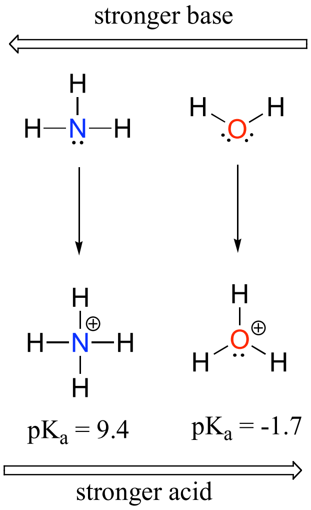
7.3A: Periodic trends

First, we will focus on individual atoms, and think about trends associated with the position of an element on the periodic table. We'll use as our first models the simple organic compounds ethane, methylamine, and methanol, but the concepts apply equally to more complex biomolecules, such as the side chains of alanine, lysine, and serine.



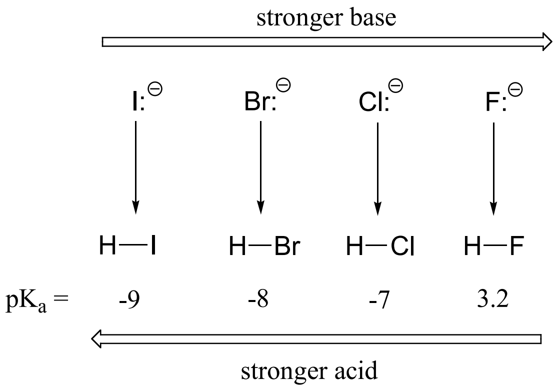
We can see a clear trend in acidity as we move from left to right along the second row of the periodic table from carbon to nitrogen to oxygen. The key to understanding this trend is to consider the hypothetical conjugate base in each case*: the more stable (weaker) the conjugate base, the stronger the acid*. Look at where the negative charge ends up in each conjugate base. In the ethyl anion, the negative charge is borne by carbon, while in the methylamine anion and methoxide anion the charges are located on a nitrogen and an oxygen, respectively. Remember the periodic trend in electronegativity (section 2.3A): it also increases as we move from left to right along a row, meaning that oxygen is the most electronegative of the three, and carbon the least. *The more electronegative an atom, the better it is able to bear a negative charge*. Thus, the methoxide anion is the most stable (lowest energy, least basic) of the three conjugate bases, and the ethyl anion is the least stable (highest energy, most basic).

We can use the same set of ideas to explain the difference in basicity between water and ammonia.



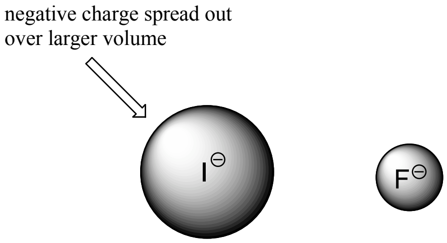
By looking at the pK_a values for the appropriate conjugate acids, we know that ammonia is more basic than water. Oxygen, as the more electronegative element, holds more tightly to its lone pair than the nitrogen. The nitrogen lone pair, therefore, is more likely to break away and form a new bond to a proton - it is, in other words, more basic. Once again, a more reactive (stronger) conjugate base means a less reactive (weaker) conjugate acid.

When moving vertically within a given column of the periodic table, we again observe a clear periodic trend in acidity. This is best illustrated with the halides: basicity, like electronegativity, increases as we move up the column.



Conversely, acidity in the haloacids increases as we move *down* the column.

In order to make sense of this trend, we will once again consider the stability of the conjugate bases. Because fluorine is the most electronegative halogen element, we might expect fluoride to also be the least basic halogen ion. But in fact, it is the *least* stable, and the most basic! It turns out that when moving vertically in the periodic table, the *size* of the atom trumps its electronegativity with regard to basicity. The atomic radius of iodine is approximately twice that of fluorine, so in an iodine ion, the negative charge is spread out over a significantly larger volume:



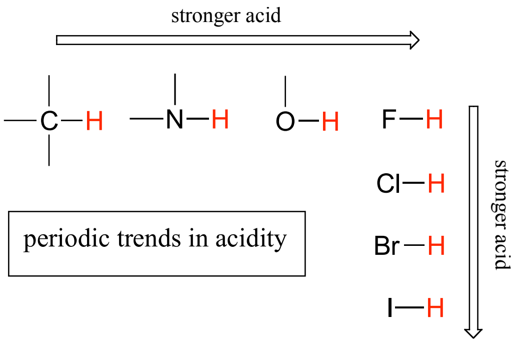
This illustrates a fundamental concept in organic chemistry that is important enough to put in its own box:

Electrostatic charges, whether positive or negative, are more stable when they are 'spread out' than when they are confined to one location.

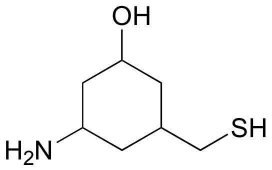
We will see this idea expressed again and again throughout our study of organic reactivity, in many different contexts. For now, the concept is applied only to the influence of atomic radius on anion stability. Because fluoride is the least stable (most basic) of the halide conjugate bases, HF is the least acidic of the haloacids, only slightly stronger than acetic acid. HI, with a pK_a of about -9, is one the strongest acids known.

More importantly to the study of biological organic chemistry, this trend tells us that thiols are more acidic than alcohols. The pK_a of the thiol group on the cysteine side chain, for example, is approximately 8.3, while the pK_a for the hydroxl on the serine side chain is on the order of 17.

To reiterate: acid strength increases as we move to the right along a row of the periodic table, and as we move down a column.



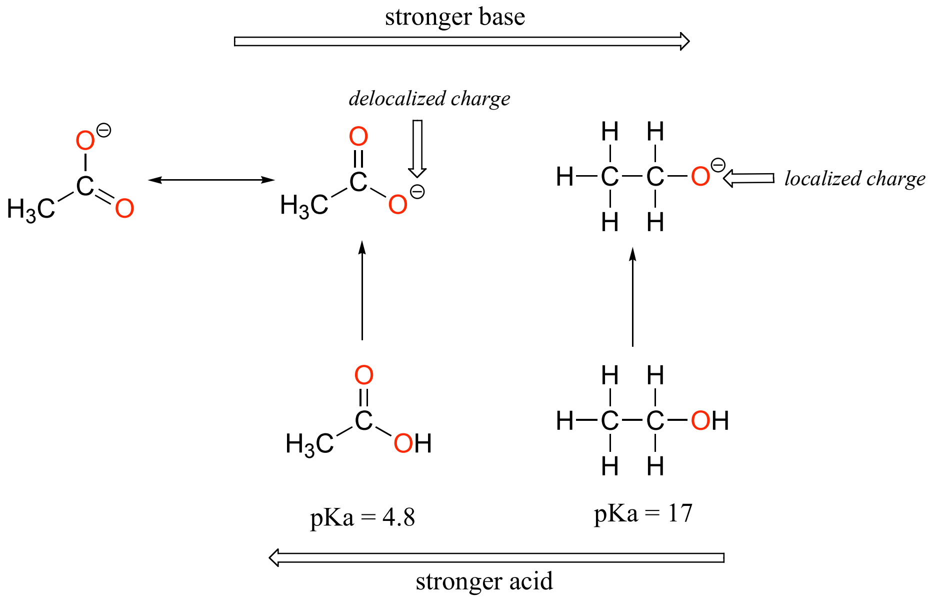
Exercise 7.6: Draw the structure of the conjugate base that would form if the compound below were to react with 1 molar equivalent of sodium hydroxide:



7.3B: The resonance effect

In the previous section we focused our attention on periodic trends - the differences in acidity and basicity between groups where the exchangeable proton was bound to different elements. Now, it is time to think about how the structure of different organic groups contributes to their relative acidity or basicity, even when we are talking about the same element acting as the proton donor/acceptor. The first model pair we will consider is ethanol and acetic acid, but the conclusions we reach will be equally valid for all alcohol and carboxylic acid groups.

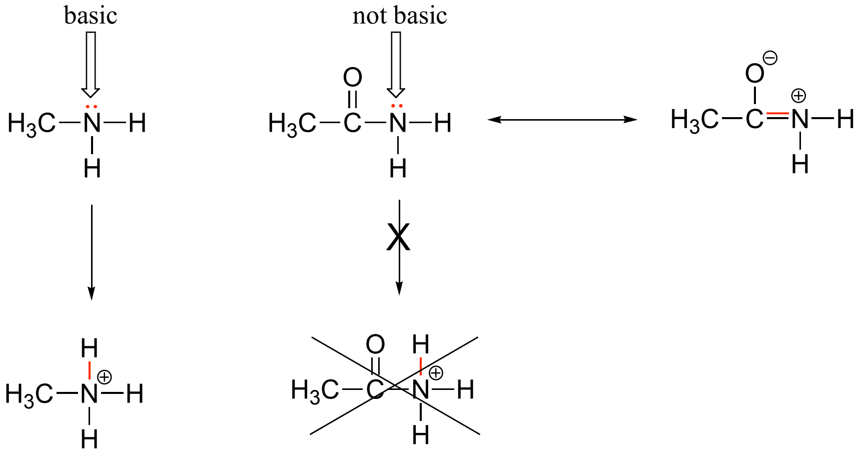
Despite the fact that they are both oxygen acids, the pK_a values of ethanol and acetic acid are very different. What makes a carboxylic acid so much more acidic than an alcohol? As before, we begin by considering the conjugate bases.



In both species, the negative charge on the conjugate base is held by an oxygen, so periodic trends cannot be invoked. For acetic acid, however, there is a key difference: a resonance contributor can be drawn in which the negative charge is localized on the second oxygen of the group. The two resonance forms for the conjugate base are equal in energy, according to our 'rules of resonance' (section 2.2C). What this means, you may recall, is that the negative charge on the acetate ion is not located on one oxygen or the other: rather it is shared between the two. Chemists use the term 'delocalization of charge' to describe this situation. In the ethoxide ion, by contrast, the negative charge is 'locked' on the single oxygen – it has nowhere else to go.

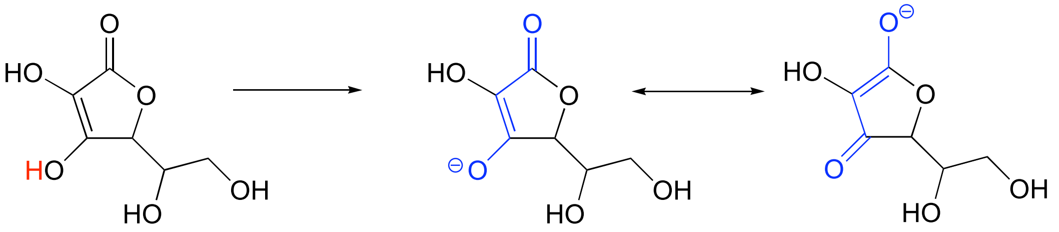
Now is the time to think back to that statement from the previous section that was so important that it got printed in bold font in its own paragraph – in fact, it is so important that we'll just say it again: "Electrostatic charges, whether positive or negative, are more stable when they are 'spread out' than when they are confined to one location." Now, we are seeing this concept in another context, where a charge is being 'spread out' (in other words, delocalized) *by resonance*, rather than simply by the size of the atom involved. The delocalization of charge by resonance has a very powerful effect on the reactivity of organic molecules, enough to account for the difference of over 12 pK_a units between ethanol and acetic acid (and remember, pK_a is a log expression, so we are talking about a difference of over 10^{12} between the acidity constants for the two molecules). The acetate ion is that much more stable than the ethoxide ion, all due to the effects of resonance delocalization.

The resonance effect also nicely explains why a nitrogen atom is basic when it is in an amine, but *not* basic when it is part of an amide group. Recall that in an amide, there is significant double-bond character to the carbon-nitrogen bond, due to a second resonance contributor in which the nitrogen lone pair is part of a π bond.



While the electron lone pair of an amine nitrogen is ‘stuck’ in one place, the lone pair on an amide nitrogen is delocalized by resonance. Notice that in this case, we are extending our central statement to say that electron density – in the form of a lone pair – is stabilized by resonance delocalization, even though there is not a negative charge involved. Here’s another way to think about it: the lone pair on an amide nitrogen is not available for bonding with a proton – these two electrons are too ‘comfortable’ being part of the delocalized π-bonding system. The lone pair on an amine nitrogen, by contrast, is not part of a delocalized π system, and is very ready to form a bond with any acidic proton that might be nearby.

Often it requires some careful thought to predict the most acidic proton on a molecule. Ascorbic acid, also known as Vitamin C, has a pK_a of 4.1.

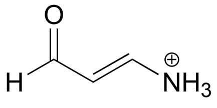


ascorbic acid
(Vitamin C)

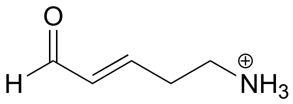
ascorbate

There are four hydroxyl groups on this molecule – which one is the most acidic? If we consider all four possible conjugate bases, we find that there is only one for which we can delocalized the negative charge over *two* oxygen atoms.

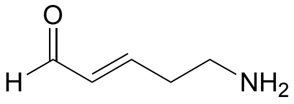
Exercise 7.7: Rank the compounds below from most acidic to least acidic, and explain your reasoning.



A



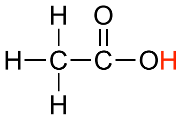
B



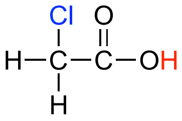
C

7.3C: The inductive effect

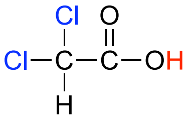
Compare the pK_a values of acetic acid and its mono-, di-, and tri-chlorinated derivatives:



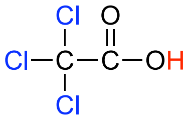
$pK_a = 4.8$



$pK_a = 2.8$



$pK_a = 1.3$

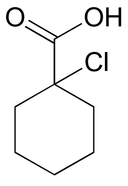


$pK_a = 0.64$

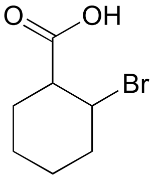
The presence of the chlorines clearly increases the acidity of the carboxylic acid group, but the argument here does not have to do with resonance delocalization, because no additional resonance contributors can be drawn for the chlorinated molecules. Rather, the explanation for this phenomenon involves something called the **inductive effect**. A chlorine atom is more electronegative than a hydrogen, and thus is able to 'induce', or 'pull' electron density towards itself, away from the carboxylate group. In effect, the chlorine atoms are helping to further spread out the electron density of the conjugate base, which as we know has a stabilizing effect. In this context, the chlorine substituent is called an **electron-withdrawing group**. Notice that the pK_a-lowering effect of each chlorine atom, while significant, is not as dramatic as the delocalizing resonance effect illustrated by the difference in pK_a values between an alcohol and a carboxylic acid. In general, *resonance effects are much more powerful than inductive effects*.

The inductive electron-withdrawing effect of the chlorines takes place through covalent bonds, and its influence decreases markedly with distance – thus a chlorine two carbons away from a carboxylic acid group has a decreased effect compared to a chlorine just one carbon away.

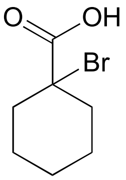
Exercise 7.8: Rank the compounds below from most acidic to least acidic, and explain your reasoning.



A



B

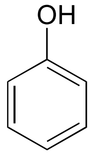


C

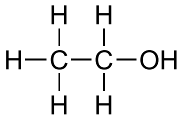
Section 7.4: More on resonance effects on acidity and basicity

7.4A: The acidity of phenols

Resonance effects involving aromatic structures can have a dramatic influence on acidity and basicity. Notice, for example, the difference in acidity between phenol and ethanol.



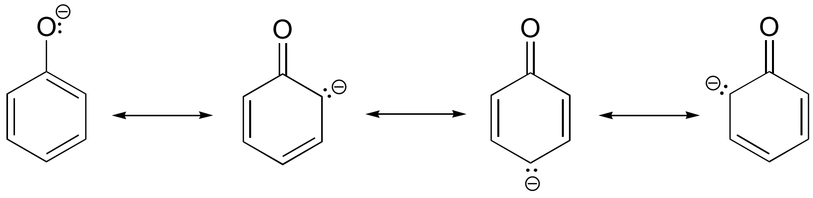
$pK_a = 10$



$$\begin{array}{ccccc} & H & & H & \\ & | & & | & \\ H- & C & - & C & -OH \\ & | & & | & \\ & H & & H & \end{array}$$

$pK_a = 17$

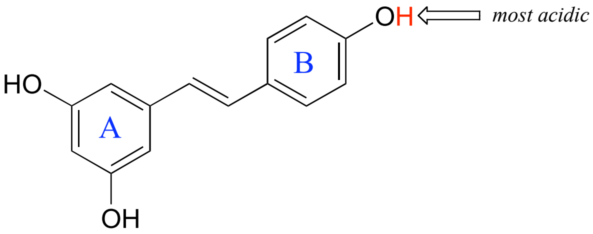
Looking at the conjugate base of phenol, we see that the negative charge can be delocalized by resonance to three different carbons on the aromatic ring.



Although these are all minor resonance contributors (negative charge is placed on a carbon rather than an oxygen), they nonetheless have a significant effect on the acidity of the phenolic proton. Essentially, the benzene ring is acting as an electron-withdrawing group.

As we begin to study in detail the mechanisms of biological organic reactions, we'll see that the phenol side chain of the amino acid tyrosine, with its relatively acidic pK_a of 9-10, often acts as a catalytic proton donor/acceptor in enzyme active sites.

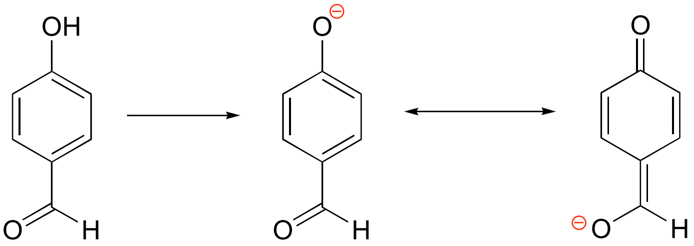
Delocalization of electron density has a stabilizing effect, and *the greater area over which the delocalization is possible, the greater the stabilization*. The most acidic of the three phenolic groups in resveretrol (an antioxidant compound in red wine) is indicated with an arrow in the figure below. This proton is expected to have a pK_a value somewhat lower than that of phenol.



We can predict this because, when we draw all the possible resonance contributors for the conjugate base, we see that the negative charge can be delocalized over the entire π structure of the molecule, to the A ring as well as to the B ring. A negative charge that results from deprotonation of either of the other two phenol groups, however, can be delocalized only over the A ring.

Exercise 7.9: Indicate, on a structural drawing of the conjugate base of resveretrol, all of the atoms to which the negative charge can be delocalized by resonance.

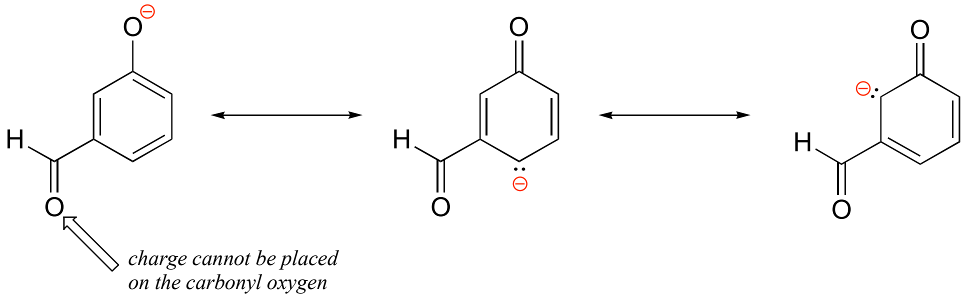
The base-stabilizing effect of an aromatic ring can also be accentuated by the presence of an additional electron-withdrawing substituent, such as a carbonyl. For the conjugate base of the phenol derivative below, an additional resonance contributor can be drawn in which the negative formal charge is placed on the carbonyl oxygen.



Now the negative charge on the conjugate base can be spread out over *two* oxygens (in addition to three aromatic carbons). The aromatic aldehyde above is expected to be substantially more acidic than phenol.

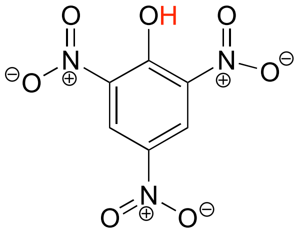
The position of the electron-withdrawing substituent relative to the phenol hydroxide is very important. When a carbonyl is in the ***para*** position (three carbons away) or the ***ortho*** position (one carbon away), it can act as an electron-withdrawing group by resonance.

In the ***meta*** position, however (two carbons away), we cannot draw a resonance contributor in which the negative formal charge is placed on the carbonyl oxygen.



In the *meta* position, the carbonyl still acts as an electron-withdrawing group, *but only by induction, not by resonance*. As stated earlier, inductive effects are generally much weaker than resonance effects.

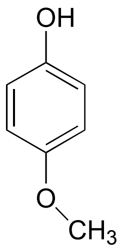
In the laboratory, nitro groups, which are very powerfully electron-withdrawing, are often used to increase the acidity of phenol groups. Picric acid, for example, has a pK_a of 0.25 – it is more acidic than trifluoroacetic acid.



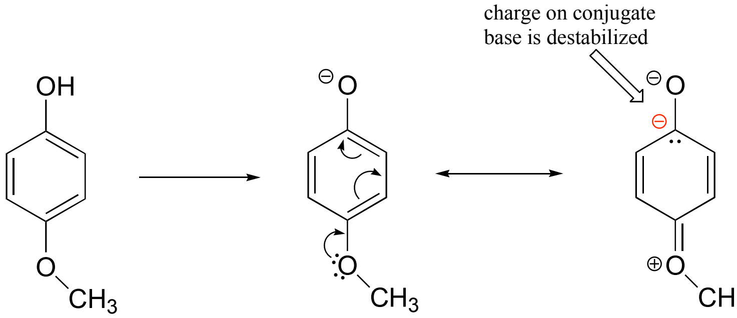
picric acid
$pK_a = 0.25$

Exercise 7.10: Use a resonance argument to explain why picric acid has such a low pK_a.

A methoxy group located at the para or ortho position of phenol causes a decrease in acidity. Consider the example of *para*-methoxyphenol.



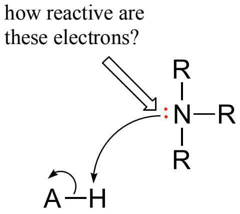
At first inspection, you might assume that the methoxy substituent, with its electronegative oxygen, is a electron-withdrawing group by induction. That is correct, but only to a point. The oxygen atom does indeed exert an electron-withdrawing inductive effect, but the lone pairs on the oxygen cause the exact opposite effect – *the methoxy group is an* ***electron-donating group*** *by resonance*. A resonance contributor can be drawn in which a formal negative charge is placed on the carbon adjacent to the negatively-charged phenolate oxygen.



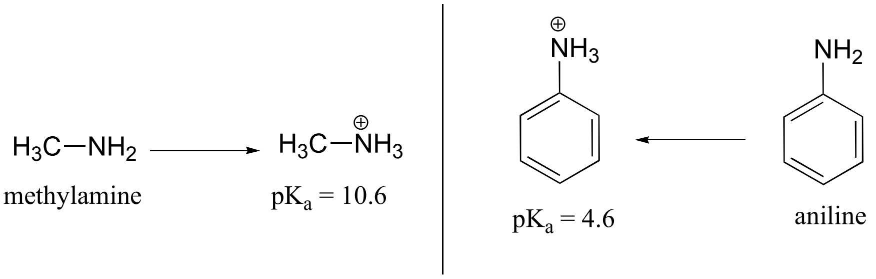
Because of like-charge repulsion, this *destabilizes* the negative charge on the phenolate oxygen, making it more reactive. It may help to visualize the methoxy group 'pushing' electrons towards the basic lone pair of the phenolate oxygen, causing them to want to reach out and grab a proton.

7.4B: The basicity of nitrogen-containing groups: aniline, imines, pyridine, and pyrrole

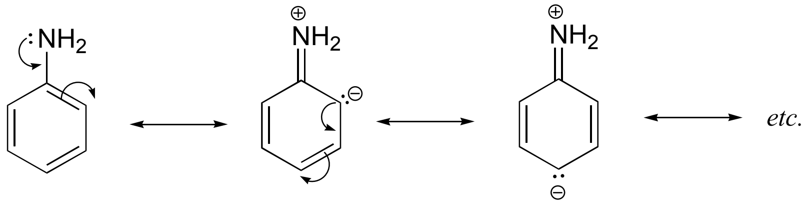
We have already learned that amines are relatively basic groups, while amides, due to resonance with the adjacent carbonyl, are not basic (see section 7.3B). Now, let's consider some other nitrogen-containing molecules in which aromatic groups exert an influence on basicity. When evaluating the basicity of a nitrogen-containing organic functional group, the central question we need to ask ourselves is: how reactive (and thus how basic) is the lone pair on the nitrogen? In other words, how much does that lone pair want to break away from the nitrogen nucleus and form a new bond with a hydrogen?



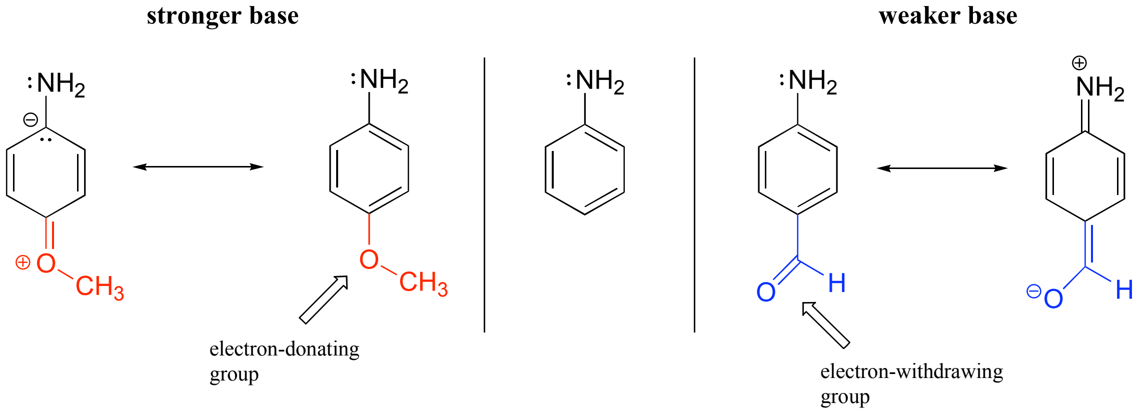
Aniline is substantially less basic than methylamine, as is evident by looking at the pK_a values for their respective ammonium conjugate acids:



This difference is basicity can be explained by the observation that, in aniline, the basic lone pair on the nitrogen is to some extent tied up in – and stabilized by – the aromatic π system.



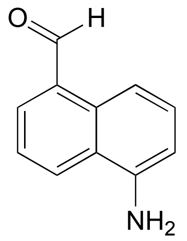
This effect is accentuated by the addition of an electron-withdrawing group such as a carbonyl, and reversed to some extent by the addition of an electron-donating group such as methoxide.



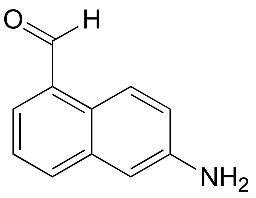
In the case of 4-methoxy aniline (the molecule on the left side of the figure above), the lone pair on the methoxy group donates electron density to the aromatic system, and a resonance contributor can be drawn in which a negative charge is placed on the carbon adjacent to the nitrogen, which makes the lone pair of the nitrogen more reactive. In effect, the methoxy group is 'pushing' electron density towards the nitrogen. Conversely, the aldehyde group on the right-side molecule is 'pulling' electron density away from the nitrogen, decreasing its basicity.

At this point, you should draw resonance structures to convince yourself that these resonance effects are possible when the substituent in question (methoxy or carbonyl) is located at the *ortho* or *para* position, but not at the *meta* position.

Exercise 7.11: Which of the two aromatic aniline-like amines below is expected to be more basic? Explain your reasoning.



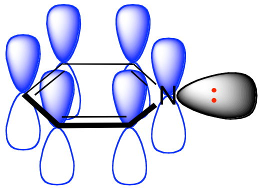
A

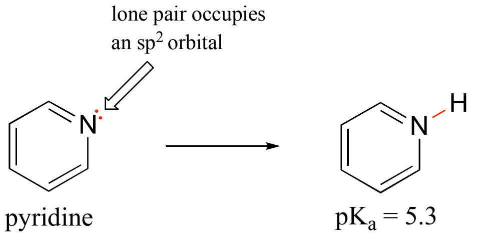


B

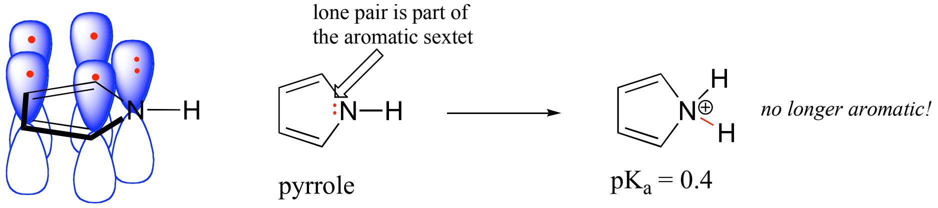
Recall from section 1.4A that an imine functional group is characterized by an sp^2-hybridized nitrogen double-bonded to a carbon. Imines are somewhat basic, with pK_a values for the protonated forms ranging around 7. Notice that this is significantly less basic than amine groups (eg. pK_a = 10.6 for methylammonium), in which the nitrogen is sp^3-hybridized. This phenomenon can be explained using orbital theory and the inductive effect: the sp^2 orbitals of an imine nitrogen are one part *s* and two parts *p*, meaning that they have about 67% *s* character. The sp^3 orbitals of an amine nitrogen, conversely, are only 25% *s* character (one part *s*, three parts *p*). Because the *s* atomic orbital holds electrons in a spherical shape, closer to the nucleus than a *p* orbital (section 1.1A), sp^2 hybridization implies greater electronegative than sp^3 hybridization. Finally, recall the inductive effect from section 7.3C: more electronegative atoms absorb electron density more easily, and thus are more acidic. Moral of the story: protonated imine nitrogens are more acidic than protonated amines, thus imines are less basic than amines.

When a nitrogen atom is incorporated directly into an aromatic ring, its basicity depends on the bonding context. In a pyridine ring, for example, the nitrogen lone pair occupies an sp^2-hybrid orbital, and is *not* part of the aromatic sextet - it is essentially an imine nitrogen. Its electron pair is available for forming a bond to a proton, and thus the pyridine nitrogen atom is somewhat basic.

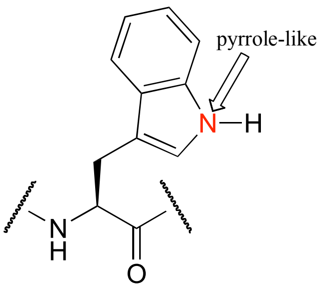




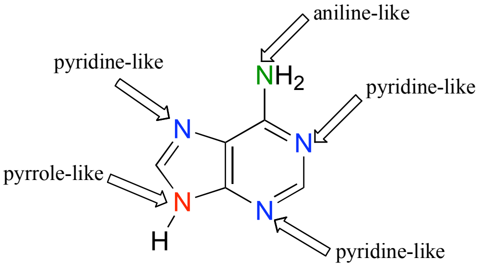
In a pyrrole ring, in contrast, the nitrogen lone pair *is* part of the aromatic sextet. This means that these electrons are very stable right where they are (in the aromatic system), and are much less available for bonding to a proton (and if they *do* pick up a proton, the aromic system is destroyed). For these reasons, pyrrole nitrogens are not strongly basic.



The aniline, pyridine, and pyrrole examples are good models for predicting the reactivity of nitrogen atoms in more complex ring systems (a huge diversity of which are found in nature). The tryptophan side chain, for example, contains a non-basic 'pyrrole-like' nitrogen, while adenine (a DNA/RNA base) contains all three types.



tryptophan

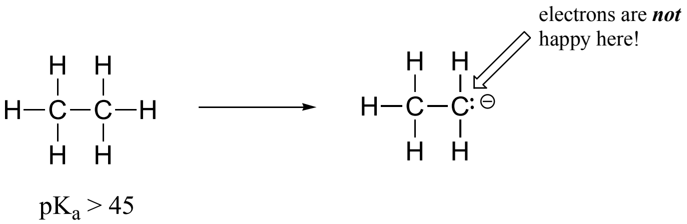


adenine

Section 7.5: Carbon acids and enolate ions

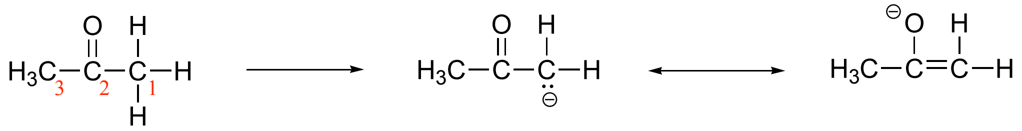
So far, we have limited our discussion of acidity and basicity to heteroatom acids, where the acidic proton is bound to an oxygen, nitrogen, sulfur, or halogen. Another very important application of acid-base theory is to the study of **carbon acids**, in which the acidic proton is bound to a carbon.

A hydrogen on an alkane is not at all acidic – its pK_a is somewhere on the order of 50, about as non-acidic as it gets in the organic chemistry world. The reason for this is that if the hydrogen were to be abstracted, the electrons from that carbon-hydrogen bond would have nowhere to go, and would have no choice but to sit uncomfortably on a single carbon.



Because carbon is not electronegative and is terrible at holding a negative charge, such carbanion species are extremely high in energy, and their formation is highly unlikely.

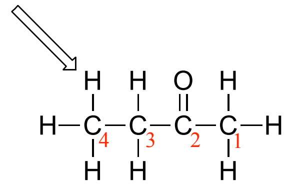
How, then, can we cause a carbon-hydrogen bond to be acidic? The answer is simple – we need to find a way to stabilize the negative charge on the carbanion conjugate base. How do we stabilize a negative charge? By spreading it out – in other words, by resonance delocalization, preferably to an electronegative atom like oxygen. This is possible when the carbon in question is located adjacent to a carbonyl group. Consider, for example, the conjugate base of acetone.



acetone
$pK_a = 19$

enolate of acetone

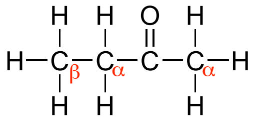
One resonance contributor puts the negative charge on the C_1 carbon . Due to the presence of the carbonyl group at C_2, however, a second resonance contributor can be drawn in which the negative charge is located on the carbonyl oxygen – where it is much happier! This type of anion species is specifically referred to as an **enolate**. The negative charge on the conjugate base is stabilized because it is delocalized, by resonance, onto an oxygen atom. Consequently, acetone is acidic –weakly so, but acidic nonetheless, with a pK_a of about 19. The importance of the position of the carbonyl group is evident when we consider 2-butanone: here, the protons on C_1 and C_3 are somewhat acidic (in the neighborhood of $pK_a = 19$), but the protons on C_4 are not acidic at all, because C_4 is not adjacent to the carbonyl.



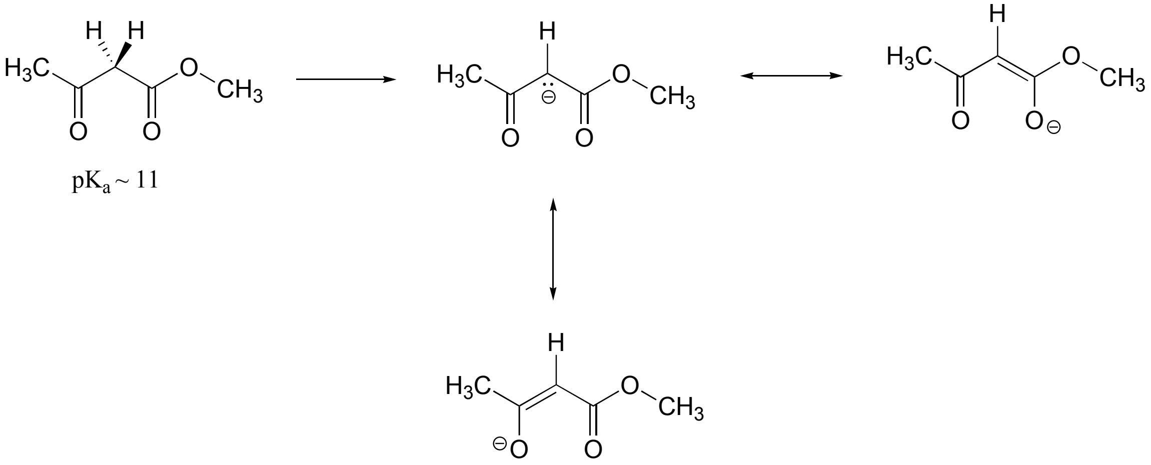
C_4 protons are not acidic

2-butanone

A carbon that is located next to a carbonyl group is referred to as an **α-carbon**, and any proton bound to it is an **α-proton**. In 2-butanone, C_1 and C_3 are α-carbons, and their five protons are α-protons. The C_4 carbon is called a β-carbon, and its three protons are β-protons.



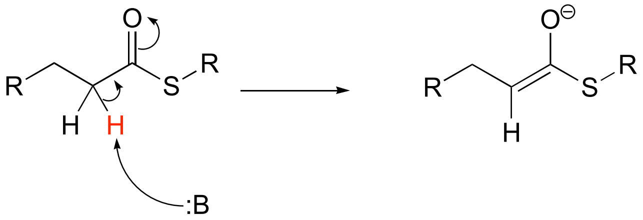
If a carbon is in the α position relative to *two* carbonyl groups rather than just one, the acidity is increased even more because of the more extensive charge delocalization that is possible for the conjugate base. This keto-ester compound, for example, has a pK_a of approximately 11, close to that of ammonium and phenol groups.



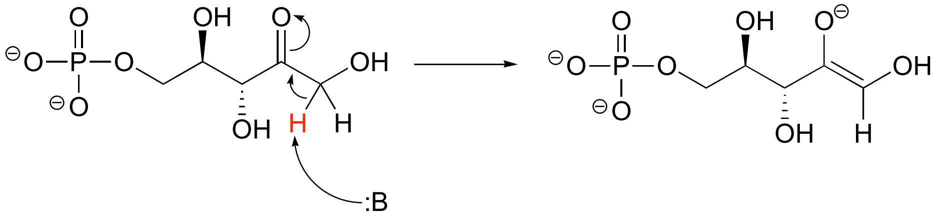
The negative charge on the conjugate base can be spread to *two* oxygens.

The acidity of α-protons is an extremely important concept in organic chemistry, biological organic chemistry in particular. Look through a biochemistry textbook, and you will see reaction after reaction in which the first mechanistic step is the abstraction of an α-proton to form an enolate intermediate. Three examples are shown here. Reaction A below is from fatty acid oxidation, while reactions B and C are both part of carbohydrate metabolism.

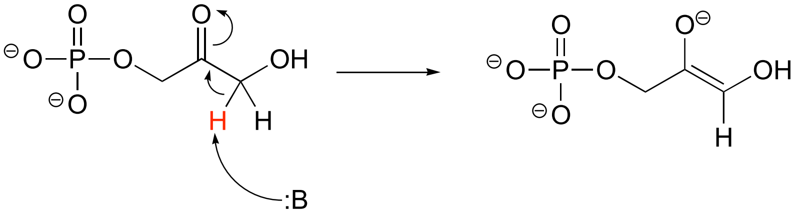
A:



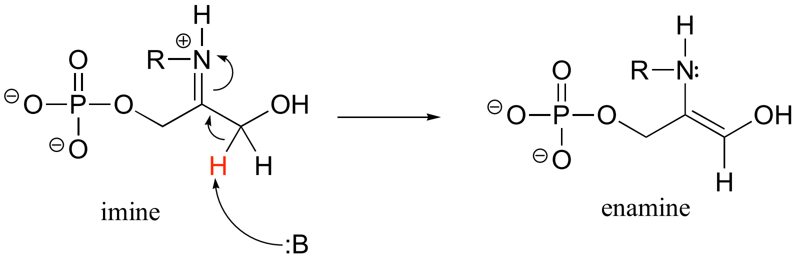
B:



C:



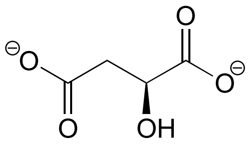
Protons that are positioned α to imine (Schiff base) groups are also acidic, due to charge delocalization onto the imine nitrogen. The conjugate base in this case is called an **enamine**.

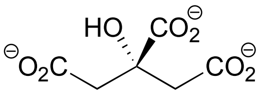


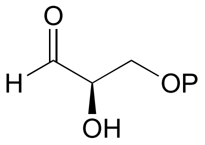
Two entire chapters in this book (13 and 14) are devoted to reactions in which a key step is the abstraction of an α-proton next to a carbonyl or imine group.

Exercise 7.12: For each molecule shown below:

a) Show the location of all α-protons.
b) Draw the structure(s) of all possible enolate conjugate bases.


malate


citrate


glyceraldehyde-3-phosphate

Terminal alkynes are another kind of carbon acid which are relevant more to laboratory organic chemistry than to biological chemistry.

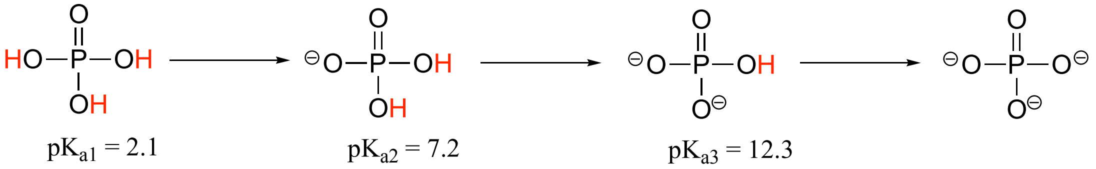
$$H_3C-C\equiv C-H \longrightarrow H_3C-C\equiv C:^{\ominus}$$

pKa ~ 26

Terminal alkynes are more acidic than alkenes or alkanes for the same reason that protonated imines are more acidic than protonated amines (section 7.4B): the alkyne carbon is sp-hybridized, meaning that it has 50% s-orbital character and is therefore more electronegative. With a pK_a of approximately 26, alkynes are only weakly acidic, but nonetheless can be fully deprotonated through the use of a strong base such as sodium amide ($NaNH_2$). As we will see in section 13.6A, this property of terminal alkynes can be useful in the organic laboratory when trying to form new carbon-carbon bonds.

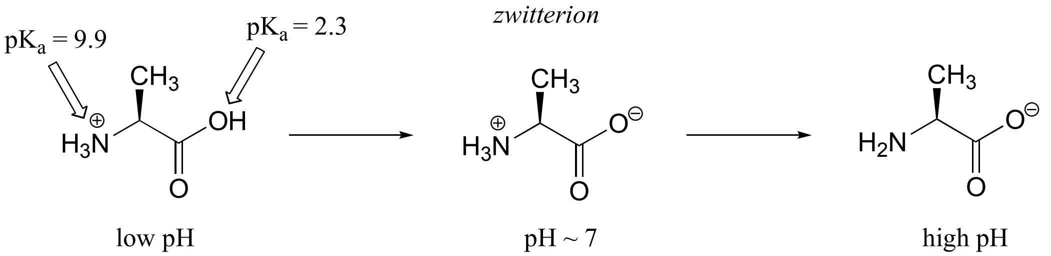
Section 7.6: Polyprotic acids

Polyprotic acids are capable of donating more than one proton. The most important polyprotic acid group from a biological standpoint is the triprotic phosphoric acid. Because phosphoric acid has three acidic protons, it also has three pK_a values.



The pK_a values for any polyprotic acid always get progressively higher, because it becomes increasingly difficult to stabilize the additional electron density that results from each successive proton donation. H_3PO_4 is a strong acid because the (single) negative charge on its conjugate base $H_2PO_4^-$ can be delocalized over two oxygen atoms. $H_2PO_4^-$ is substantially less acidic, because proton donation now results in the formation of an additional negative charge, and a –2 charge is inherently higher in energy than a –1 charge, due to negative-negative electrostatic repulsion. The third deprotonation, with formation of a third negative charge, is harder still. We will have more to say about the acidity of phosphate groups in chapter 10.

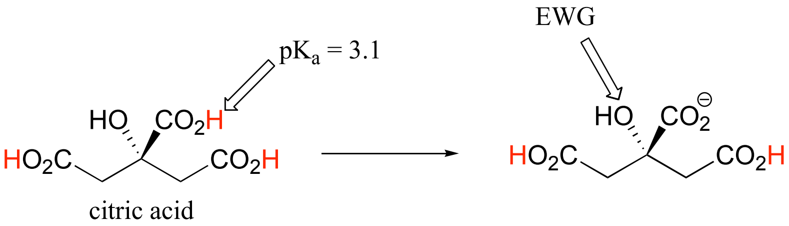
Free amino acids are polyprotic, with pK_a values of approximately 2 for the carboxylic acid group and 9-10 for the amino group. Alanine, for example, has the acid constants $pK_{a1} = 2.3$ and $pK_{a2} = 9.9$.



Alanine is almost fully protonated and positively charged when dissolved in a solution that is buffered to pH 0.5. At pH 7 , alanine has lost one proton from the carboxylic acid group, and thus is has both a negative and a positive charged group – it is now referred to as a **zwitterion**. At pH levels above 12, the ammonium group is almost fully deprotonated, and alanine has a negative overall charge.

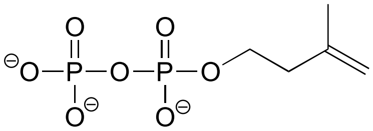
Some amino acids (arginine, lysine, aspartate, glutamate, tyrosine, and histidine) are triprotic, with a third pK_a value associated with the ionizable functional group on the side chain.

Many biological organic molecules have several potentially ionizable functional groups and thus can be considered polyprotic acids. Citric acid, found in abundance in oranges, lemons, and other citrus fruits, has three carboxylic acid groups and pK_a values of 3.1, 4.8, and 6.4.



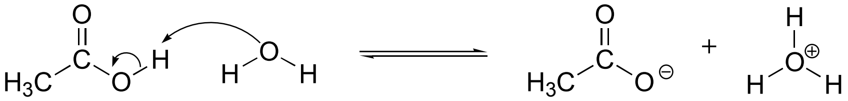
Careful consideration of the structure of citric acid can allow us to predict that it is the middle acid group that has the lowest pK_a. This is because the hydroxyl group also bonded to the middle carbon is electron-withdrawing *by induction*, and a negative charge associated with a conjugate base will be stabilized to the greatest extent on the middle carboxylate. (Notice that in this case, the oxygen, due to its position in the molecule, is *not* able to exert an electron-donating resonance effect.)

Exercise 7.13: Predict the structure of the organic diphosphate compound below in the protonation state where it has a -2 charge.



Section 7.7: The effect of solvent and enzyme microenvironment on acidity

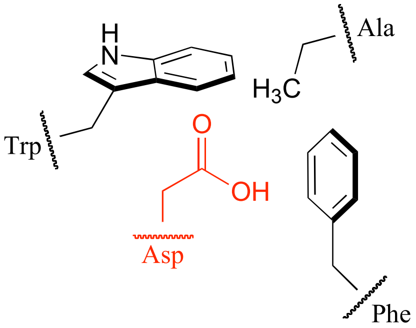
Thus far in our discussion of acidity and basicity, we have not mentioned one very important factor– the effect of the solvent. The acidity of any compound, as we well know, depends on the difference in stability between the acid and its conjugate base. The ability of a carboxylic acid to donate a proton when in aqueous solution, for example, depends in large part on how the charged products that result from proton donation are stabilized through extensive hydrogen bonding interactions with the water solvent. The importance of the role played by solvent in this reaction becomes obvious when you consider that acetic acid, when reacting with water in the gas phase, has a measured pK_a value *greater than 100*! (Recall that the pK_a of acetic acid in water is 4.8).



no solvent to stabilize charges!

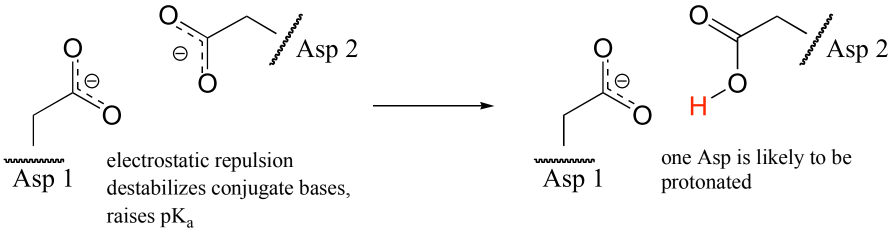
Without the solvating, ion-stabilizing effect of solvent water, it is very difficult to pull apart the acetate anion and hydronium cation.

Because water is the solvent in biological chemistry, it is tempting to assume that solvent effects are a constant factor when considering the acidity or basicity of biomolecules. This is only partially correct. Keep in mind that virtually all biochemical reactions take place inside the active site pocket of an enzyme, rather than free in solution. The actual '**microenvironment'** of an enzyme active site can often be very different from the environment outside in the 'bulk solvent'. Consider, for example, the side chain carboxylate on an aspartate residue of an enzyme. The literature pK_a of this carboxylic acid group is listed as 3.9, but this estimate assumes that the group is on the surface of the protein, exposed to solvent water. In the physiological buffer of pH ~ 7, a carboxylic acid group with pK_a = 3.9 will be fully deprotonated and negatively charged. If, however, an aspartate side chain is buried deep inside the interior of the protein's active site, and is surrounded primarily by hydrophobic residues such as alanine, phenylalanine, tryptophan, etc., the situation is very different.



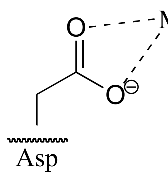
Cut off from the environment of the bulk solvent, the carboxylate group is now in a very hydrophobic microenvironment, a situation in which the protonated, uncharged state is stabilized relative to the deprotonated, negatively charged state. The overall effect is that the pK_a for this aspartate residue is higher than 3.9 – it is less acidic, and more likely to be in its protonated form inside the protein.

A similar effect would be observed in a situation where the side chain carboxylate groups of two aspartate residues are located in close proximity to one another. Two negatively charged groups close to each other represents a very high energy, repulsive situation, and this can be relieved if one of the two side chains is protonated.


O
O
Asp 2
Asp 1
electrostatic repulsion destabilizes conjugate bases, raises pK_a
H
one Asp is likely to be protonated

In this microenvironment, the proximity of one amino acid group directly effects the pK_a of its neighbor.

Now consider a situation where a metal ion such as magnesium (Mg^{+2}) or zinc (Zn^{+2}) is bound in the interior of the enzyme, in close contact with an aspartate side chain. With a cation to interact with, the anionic, deprotonated state of the amino acid is stabilized, so the pK_a of this Asp residue is likely to be substantially *lower* than 3.9.



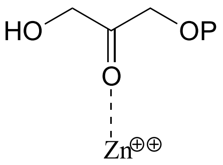
carboxylate state stabilized, $pK_a < 3.9$

The metal ion in this situation is considered to be acting as a Lewis acid, (section 7.1B), accepting electron density from the carboxylate group.

The pK_a-lowering effect of a metal cation can be dramatic – it has been estimated that a water molecule coordinated to a Cu^{+2} or Zn^{+2} ion can have a pK_a as low as 7 (compare this to the 'normal' water pK_a of 15.7!) (*J. Am. Chem. Soc.* **1984**, *106*, 630.)

Exercise 7.14: A lysine residue located deep in the interior of a protein is surrounded by alanines and leucines. In what direction will this alter the 'normal' pK_a of the lysine side chain, and why?

Exercise 7.15: In many biochemical reactions which involve the formation of an enolate intermediate, the carbonyl oxygen of the substrate is coordinated to a divalent metal ion (usually zinc or magnesium) in the active site. Explain, with structural drawings, how this ion-dipole interaction effects the acidity of the α-protons of DHAP.



The microenvironment effect is very important in enzyme catalysis, because it allows an enzyme to 'tune' the pK_a of an amino acid side chain to whatever level is optimal for catalysis. Without this effect, the only residue that would be capable of acting as a catalytic acid or base would be histidine, because it is the only side chain group with an 'untuned' pK_a value near physiological pH 7. Lacking any 'tuning' by the protein microenvironment, aspartate and glutamate side chains would both be far too acidic to act as proton acceptors in an enzymatic mechanism – they would *always* be deprotonated. Lysine, arginine, tyrosine, and cysteine, on the other hand, would not be acidic enough to donate their protons. As we will see when we begin to study organic mechanisms in detail, all of these residues regularly act as proton donors and acceptors in the course of enzyme-catalyzed reactions.

Chapter 7 problems

P7.1: In an enzyme, an Asp side chain is surrounded by phenylalanine, alanine, tryptophan, and leucine residues. Another Asp side chain located on the surface of the protein, pointing out into the surrounding water. Which residue has the higher pK_a, and why?

P7.2: (a-d) How would the immediate proximity of a magnesium ion affect the pK_a of the side chains of the following amino acids (relative to the 'typical' pK_a values given in the text)? Assume that all residues are located in the interior of the protein structure, not in direct contact with the outside buffer solution.

a) a glutamate residue?

b) a lysine residue?

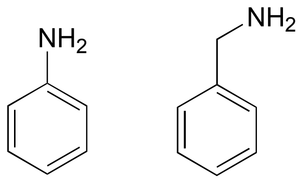
c) a histidine residue?

d) a tyrosine residue?

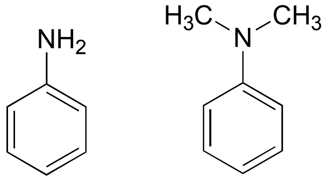
e) How would contact with a magnesium ion effect the pK_a of a bound water molecule in the interior of a protein?

P7.3: For each pair of molecules below, choose the stronger base.

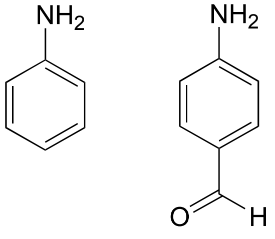
a)



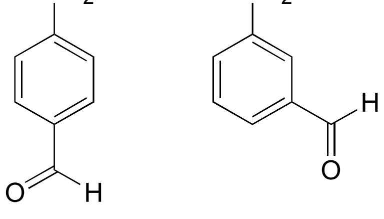
b)



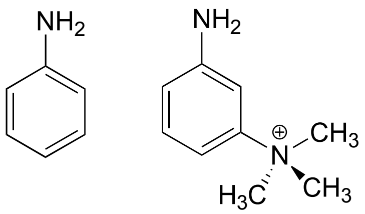
c)



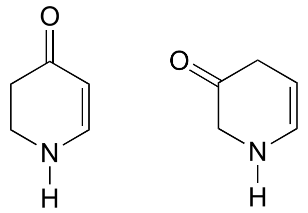
d)



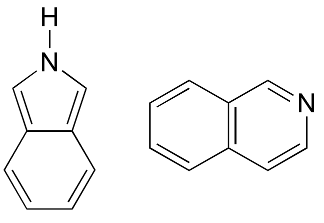
e)



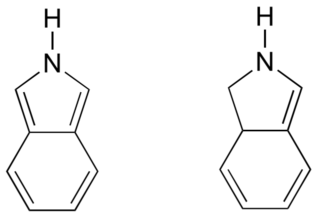
f)



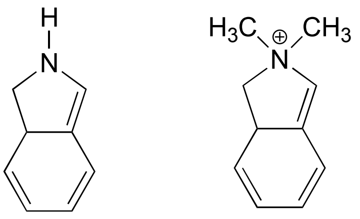
g)



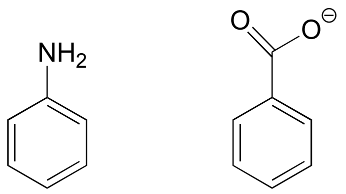
h)



i)



j)



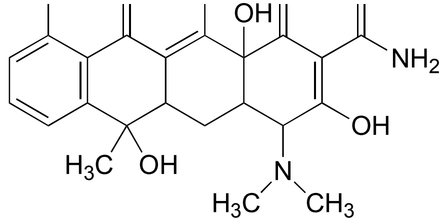
P7.4: The side chain of lysine has a pK_a of approximately 10.5, while the pK_a of the arginine side chain is approximately 12.5. Use resonance structures to rationalize this difference.

P7.5: The α-protons of ketones are, in general, significantly more acidic than those of esters. Account for this observation using structural arguments.

P7.6:

a) Locate the most acidic proton on the antibiotic tetracycline, and explain your choice.

b) Draw the structure of the conjugate base of tetracycline that has reacted with two molar equivalents of a strong base.



tetracycline

P7.7: Below are the structures of some well-known drugs.

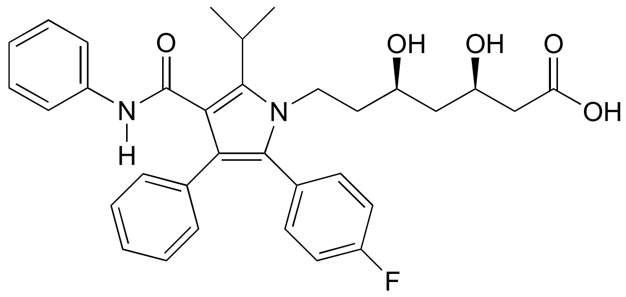
a) What is the most acidic proton on Lipitor? What is its approximate pK_a value?

b) What is the second most acidic proton on Lipitor?

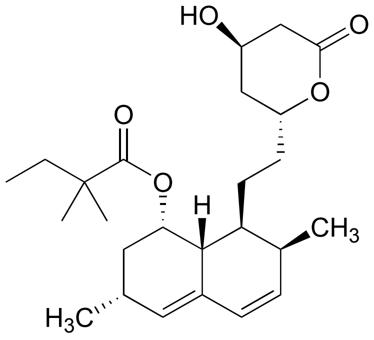
c) What is the most acidic proton on Zocor? What is its approximate pK_a value?

d) What is the most acidic proton on Plavix? What is its approximate pK_a value?

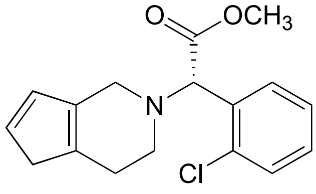
e) Where is the most basic site on Plavix?



Lipitor

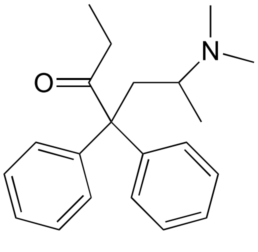


Zocor



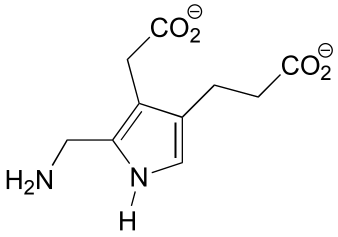
Plavix

P7.8: Draw structures for the conjugate acid and the conjugate base of methadone, an opiate used in the treatment of heroin addiction.



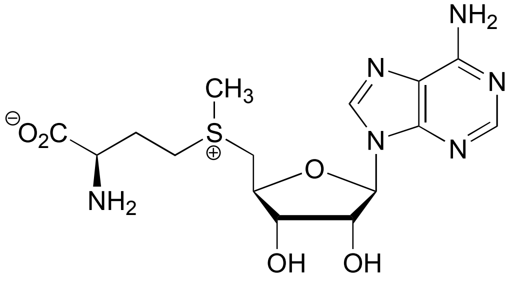
methadone

P7.9: Porphobilinogen is a precuror to the biosynthesis of chlorophyll and many other biological molecules. One of the nitrogen atoms is basic, the other is not. Which is which? Explain your reasoning.



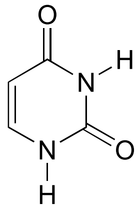
porphobilinogen

P7.10: Classify each of the amine groups in S-adenosylmethionine as an alkyl amine, aryl amine, 'pyridine-like' amine, or 'pyrrole-like' amine. Which is most basic? Which is least basic?



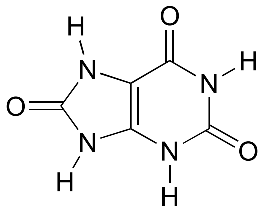
S-adenosylmethionine

P7.11: Locate the most acidic proton on the RNA base uracil, and use resonance structures of conjugate bases to explain your reasoning.



uracil

P7.12: Uric acid, an intermediate in the catabolism (breakdown) of the nucleotide adenosine, has four protons. Which is the least acidic? Use resonance structures to explain your reasoning.



uric acid

P7.13: Estimate the total charge on a peptide with the sequence P-E-P-T-I-D-E (single-letter amino acid code), when it is dissolved in a buffer with pH = 7.3 (don't forget to consider the terminal amino and carboxylate groups).

P7.14: Estimate the total charge on a dipeptide of sequence D-I.

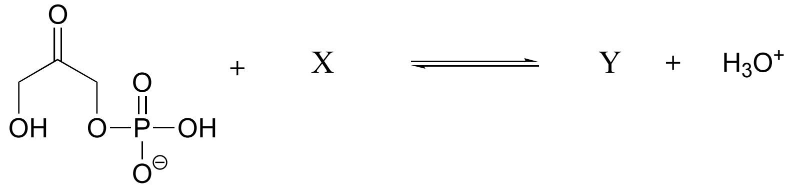
a) in a buffer with pH = 4.0

b) in a buffer with pH = 7.3

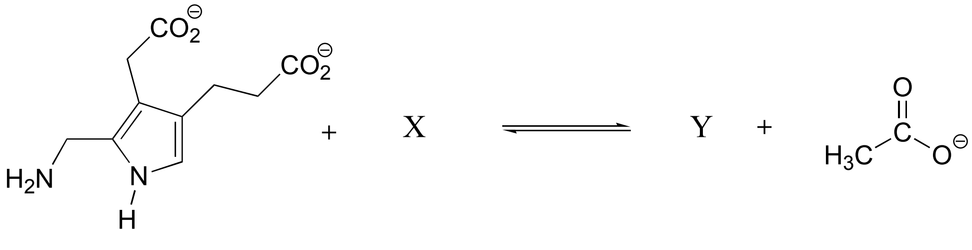
c) in a buffer with pH = 9.6

P7.15: Show the structures of species X and Y in the following acid-base reactions, and estimate the value of K_{eq} using the pK_a table. Assume that reactions involve equimolar amounts of acid and base.

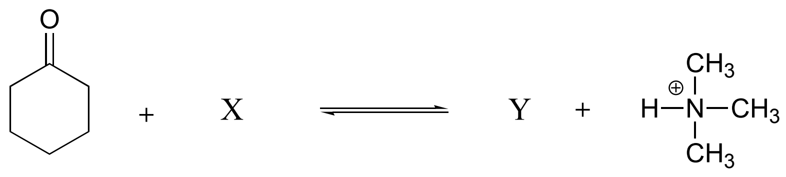
a)



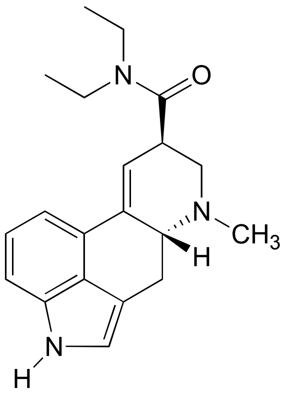
b)



c)

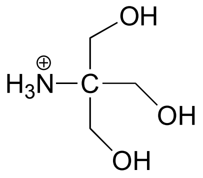


P7.16: Locate the most basic site on the structure of the hallucinogenic drug known as LSD.

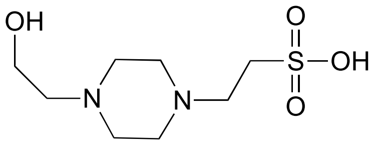


lysergic acid diethylamide
(LSD)

P7.17: Aqueous solutions of 'Tris' and 'HEPES' are very commonly used as buffers in biochemistry and molecular biology laboratories. You make two buffer solutions: One is 50 mM Tris at pH 7.0, the other 50 mM HEPES at pH 7.0. For each solution, calculate the concentration of buffer molecules that are in their charged (ionic) protonation states. *Hint* - what do you predict is this most acidic proton on HEPES?



'Tris' buffer
($pK_a = 8.1$)



'HEPES' buffer
($pK_a = 7.5$)

Chapter 8

Nucleophilic substitution reactions, part I

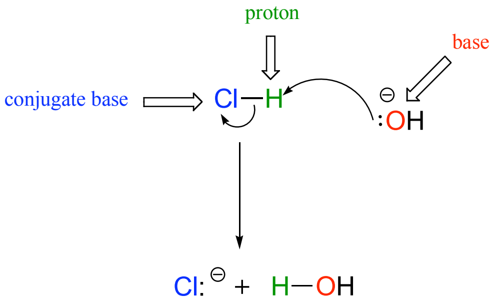
Introduction

In this chapter, we will extend our discussion, begun in chapter 6, of those organic reactions commonly known as 'nucleophilic substitutions'. In doing do, we will have an opportunity to consider in greater detail the various factors affecting the three main players in most organic reaction mechanisms: the nucleophile, the electrophile, and the leaving group. We will also see, for the first time, how stereochemical concepts are applied to the analysis of an organic reaction.

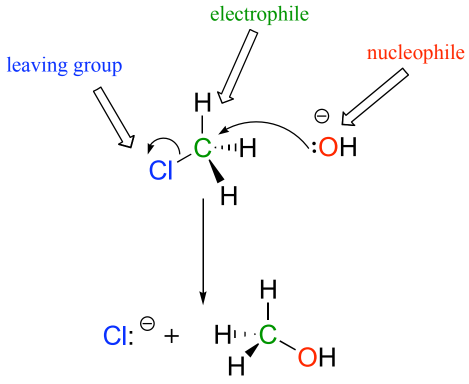
As we tackle this new and challenging chemistry, it is very important to keep in mind the central ideas that we learned in the acid-base chapter about electron density and how it is stabilized. At that time, we were talking about the breaking and forming of bonds to acidic hydrogens. Now, we will extend our discussion to reactions involving bonds between heteroatoms and carbon - but for much of what concerns us in this chapter, all of the fundamental ideas we've learned about electron density still apply!

Section 8.1: Introduction to the nucleophilic substitution reaction

Recall from chapter 6 that, in many ways, the proton transfer process in a Brønsted-Lowry acid-base reaction can be thought of as simply a special kind of nucleophilic substitution reaction, one in which the electrophile is a hydrogen rather than a carbon.



acid-base reaction

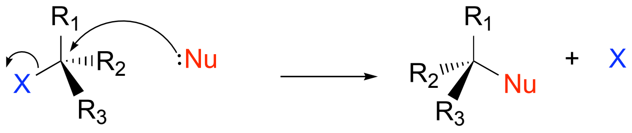


nucleophilic substitution

In both reaction types, we are looking at very similar players: an electron-rich species (the nucleophile/base) attacks an electron-poor species (the electrophile/proton), driving off the leaving group/conjugate base.

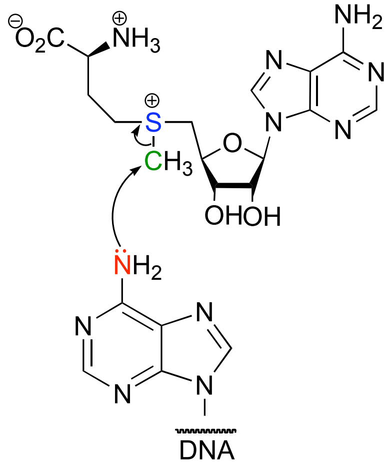
In the next few sections, we are going to be discussing some general aspects of nucleophilic substitution reactions, and in doing so it will simplify things greatly if we can use some abbreviations and generalizations before we dive into real examples. Instead of showing a specific nucleophile like hydroxide, we will simply refer to the nucleophilic reactant as 'Nu'. In a similar fashion, we will call the leaving group 'X'. We will see as we study actual reactions that leaving groups are sometimes negatively charged, sometimes neutral, and sometimes positively charged. We will also see some examples of nucleophiles that are negatively charged and some that are neutral. Therefore, in this general picture we will not include a charge designation on the 'X' or 'Nu' species. In the same way, we will see later that nucleophiles and leaving groups are sometimes protonated and sometimes not, so for now, for the sake of simplicity, we will not include protons on 'Nu' or 'X'. We will generalize the three other groups bonded on the electrophilic central carbon as R_1, R_2, and R_3: these symbols could represent hydrogens as well as alkyl groups. Finally, in order to keep figures from becoming too crowded, we will use in most cases the line structure convention in which the central, electrophilic carbon is not drawn out as a 'C'.

Here, then, is the generalized picture of a concerted (single-step) nucleophilic substitution reaction:

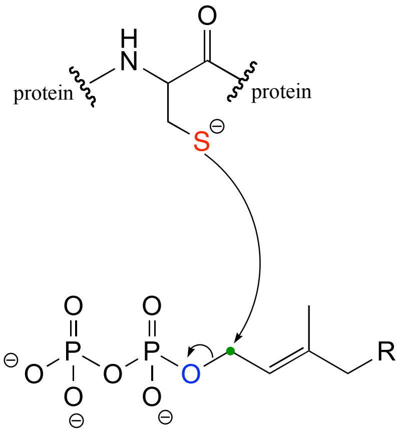


In order to provide a more realistic picture, three actual nucleophilic substitution reactions are shown below. Reactions A and B are biological reactions, while C is a laboratory synthesis reaction (reaction A should be familiar from chapter 6). In each case, the nucleophilic atom is colored red and the leaving group atom is blue. We will examine each of these reactions in more detail in the next chapter - for now, just make sure that you can relate what is happening in these actual reactions to the general picture shown above. (Reaction A: Nucleic Acids Research 2000, 28, 3950. Reaction B: Biochemistry 1998, 37, 16601. Reaction C: Carbohydrate Research 2004, 339, 51.)

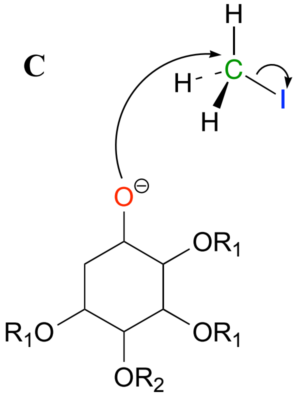
A



B



C

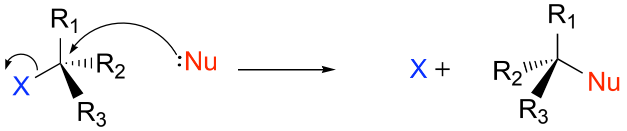


Exercise 8.1: Draw the products of the three nucleophilic substitution reactions shown above. For reactions A and C, assume that a proton is removed from the nucleophilic atom during or after nucleophilic attack.

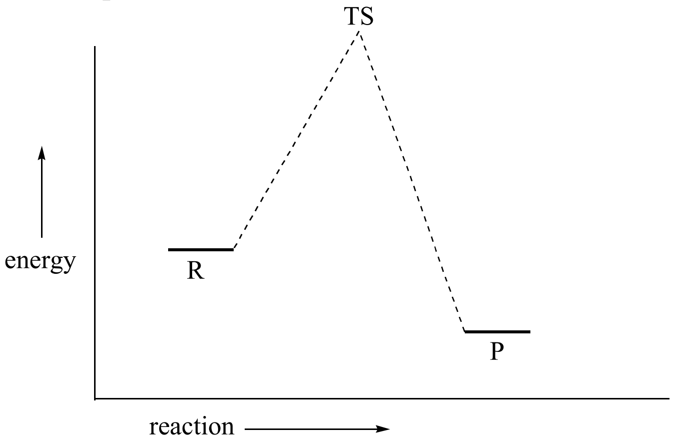
Section 8.2: Two mechanistic models for a nucleophilic substitution reaction

8.2A: Associative nucleophilic substitution: the S_N2 reaction

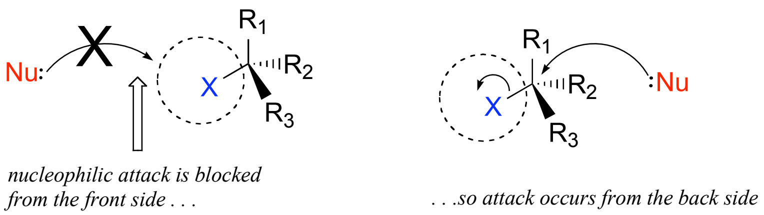
You may recall from chapter 6 that there are two mechanistic models for how a nucleophilic substitution reaction can proceed. In the first picture, the reaction takes place in a single step, and bond-forming and bond-breaking occur simultaneously.



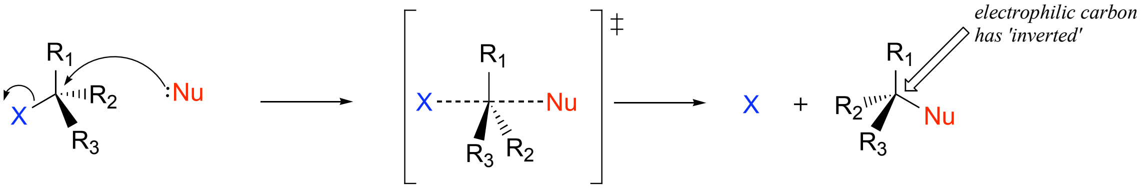
This is called an '**associative**', or '**S_N2**' mechanism. In the term S_N2, S stands for 'substitution', the subscript N stands for 'nucleophilic', and the number 2 refers to the fact that this is a **bimolecular reaction**: the overall rate depends on a step in which two separate molecules (the nucleophile and the electrophile) collide. A potential energy diagram for this reaction shows the transition state (TS) as the highest point on the pathway from reactants to products.



If you look carefully at the progress of the S_N2 reaction, you will realize something very important about the outcome. The nucleophile, being an electron-rich species, must attack the electrophilic carbon from the *back side* relative to the location of the leaving group. Approach from the front side simply doesn't work: the leaving group - which is also an electron-rich group - blocks the way.


Nu:
R1
X
R2
R3
nucleophilic attack is blocked from the front side . . .
R1
X
R2
R3
:Nu
. . .so attack occurs from the back side

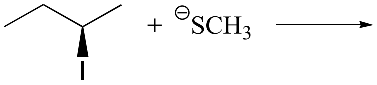
The result of this backside attack is that the stereochemical configuration at the central carbon *inverts* as the reaction proceeds. In a sense, the molecule is turned inside out. At the transition state, the electrophilic carbon and the three 'R' substituents all lie on the same plane.



transition state

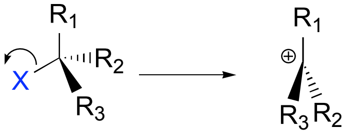
What this means is that S_N2 reactions whether enzyme catalyzed or not, are inherently stereoselective: when the substitution takes place at a stereocenter, we can confidently predict the stereochemical configuration of the product.

Exercise 8.2: Predict the structure of the product in this S_N2 reaction. Be sure to specify stereochemistry.

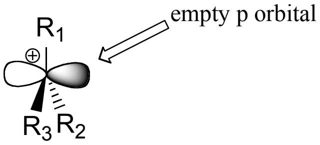


8.2B: Dissociative nucleophilic substitution: the S_N1 reaction

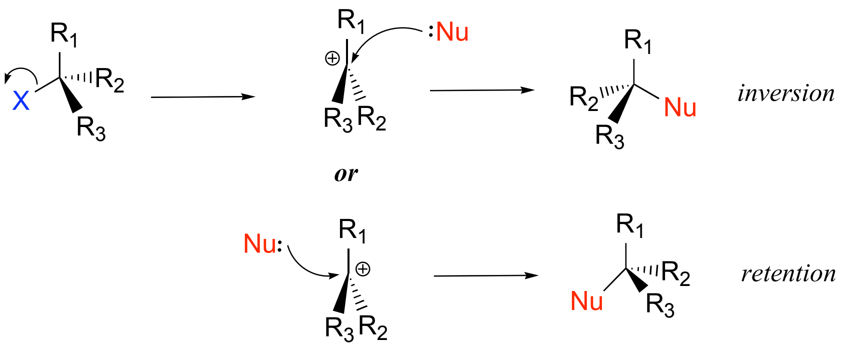
A second model for a nucleophilic substitution reaction is called the '**dissociative**', or '**S_N1**' mechanism: in this picture, the C-X bond breaks *first*, before the nucleophile approaches:



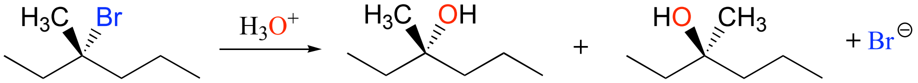
This results in the formation of a carbocation: because the central carbon has only three bonds, it bears a formal charge of +1. Recall that a carbocation should be pictured as sp^2 hybridized, with trigonal planar geometry. Perpendicular to the plane formed by the three sp^2 hybrid orbitals is an empty, unhybridized *p* orbital.



In the second step of this two-step reaction, the nucleophile attacks the empty, 'electron hungry' *p* orbital of the carbocation to form a new bond and return the carbon to tetrahedral geometry.



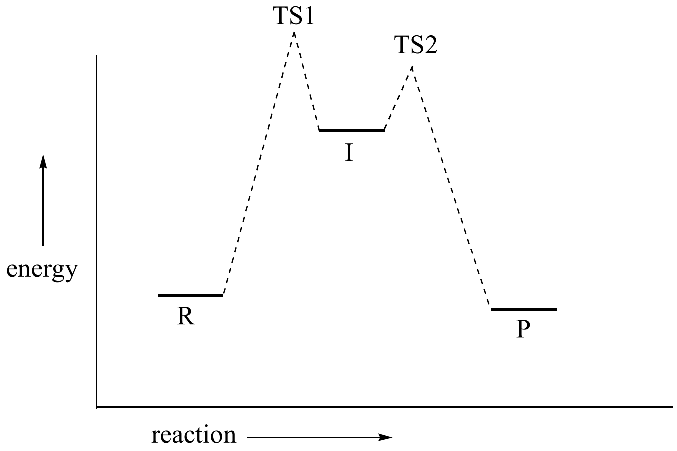
We saw that S_N2 reactions result specifically in inversion of stereochemistry at the electrophilic carbon center. What about the stereochemical outcome of S_N1 reactions? In the model S_N1 reaction shown above, the leaving group dissociates completely from the vicinity of the reaction before the nucleophile begins its attack. Because the leaving group is no longer in the picture, the nucleophile is free to attack from either side of the planar, *sp*^{2}-hybridized carbocation electrophile. This means that about half the time the product has the same stereochemical configuration as the starting material (retention of configuration), and about half the time the stereochemistry has been inverted. In other words, *racemization* has occurred at the carbon center. As an example, the tertiary alkyl bromide below would be expected to form a racemic mix of *R* and *S* alcohols after an S_N1 reaction with water as the incoming nucleophile.

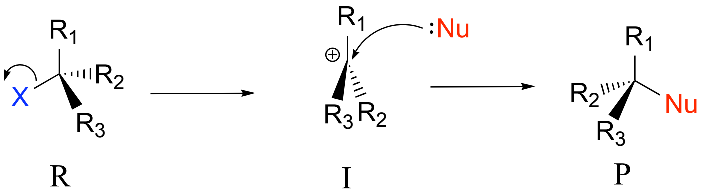


Exercise 8.3: Draw the structure of the intermediate in this two-step reaction.

While most nonenzymatic S_N1 reactions are not stereoselective, we will see later that enzyme-catalyzed nucleophilic substitution reactions - whether S_N1 or S_N2 - almost always *are* stereoselective. The direct result of an enzymatic nucleophilic substitution reaction is more often than not *inversion* of configuration - this is because the leaving group usually remains bound in the enzyme's active site long enough to block a nucleophilic attack from that side. This does not mean, however, that enzymes can only catalyze substitution reactions with inversion of configuration: we will see in the next chapter (section 9.2) an example of an enzymatic nucleophilic substitution reaction in which the overall result is 100% *retention* of configuration.

In the S_N1 reaction we see an example of a reaction intermediate, a very important concept in the study of organic reaction mechanisms that was first introduced in Chapter 6. Recall that many important organic reactions do not occur in a single step; rather, they are the sum of two or more discreet bond-forming / bond-breaking steps, and involve transient intermediate species that go on to react very quickly. In the S_N1 reaction, the carbocation species is a reaction intermediate. A potential energy diagram for an S_N1 reaction shows that the carbocation intermediate can be visualized as a kind of valley in the path of the reaction, higher in energy than both the reactant and product but lower in energy than the two transition states.





Exercise 8.4: Draw structures representing TS1 and TS2 in the reaction above. Use the solid/dash wedge convention to show three dimensions.

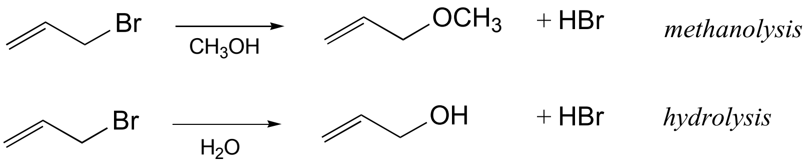
Recall (section 6.2) that the first step, in which two charged species are formed from a neutral molecule, is much the slower of the two steps, and is therefore rate-determining. This is illustrated by the energy diagram, where the activation energy for the first step is higher than that for the second step. Also recall that an S_N1 reaction has *first order* kinetics, because the rate determining step involves one molecule splitting apart, not two molecules colliding.

Exercise 8.5: Consider two nucleophilic substitutions that occur uncatalyzed in solution. Assume that reaction A is S_N2, and reaction B is S_N1. Predict, in each case, what would happen to the rate of the reaction if the concentration of the nucleophile were doubled, while all other conditions remained constant.

A: $CH_3I + CH_3S^{\ominus} \xrightarrow{S_N2}$

B: $(CH_3)_3CBr + CH_3SH \xrightarrow{S_N1}$

Many S_N1 reactions are of a class that are referred to as **solvolysis**, where a solvent molecule participates in the reaction as a nucleophile. The S_N1 reaction of allyl bromide in methanol is an example of what we would call **methanolysis**, while if water is the solvent the reaction would be called **hydrolysis**:

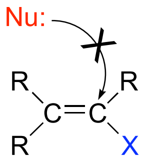


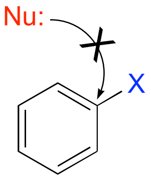
Because water and alcohols are relatively weak nucleophiles, they are unlikely to react in an S_N2 fashion.

Exercise 8.6: Draw a complete curved-arrow mechanism for the methanolysis reaction above.

8.2C: Nucleophilic substitutions occur at sp^3-hybridized carbons

One more important point must be made before continuing: nucleophilic substitutions as a rule occur at sp^3-hybridized carbons, and *not* where the leaving group is attached to an sp^2-hybridized carbon::





Bonds on sp^2-hybridized carbons are inherently shorter and stronger than bonds on sp^3-hybridized carbons (section 1.5C), meaning that it is harder to break the C-X bond in these substrates. S_N2 reactions of this type are unlikely also because the (hypothetical) electrophilic carbon is protected from nucleophilic attack by electron density in the π bond. S_N1 reactions are highly unlikely, because the resulting carbocation intermediate, which would be sp-hybridized, would be very unstable (we'll discuss the relative stability of carbocation intermediates in section 8.4B).

Before we look at some real-life nucleophilic substitution reactions in the next chapter, we will spend some time in the remainder of this chapter focusing more closely on the three principal partners in the nucleophilic substitution reaction: the nucleophile, the electrophile, and the leaving group.

Section 8.3: More about nucleophiles

8.3A: What makes a nucleophile?

As a general rule, nucleophile substitution reactions that involve powerful nucleophiles tend to occur with S_N2 mechanisms, while reactions with weaker nucleophiles tend to be S_N1. But what makes a group a strong or weak nucleophile? Nucleophilic functional groups are those which have electron-rich atoms able to donate a pair of electrons to form a new covalent bond. In both laboratory and biological organic chemistry, the most relevant nucleophilic atoms are oxygen, nitrogen, and sulfur, and the most common nucleophilic functional groups are water, alcohols, phenols, amines, thiols, and occasionally carboxylates.

More specifically in laboratory reactions, halide and azide (N_3^-) anions are commonly seen acting as nucleophiles.

Of course, carbons can also be nucleophiles - otherwise how could new carbon-carbon bonds be formed in the synthesis of large organic molecules like DNA or fatty acids? Enolate ions (section 7.5) are the most common carbon nucleophiles in biochemical reactions, while the cyanide ion (CN^-) is just one example of a carbon nucleophile commonly used in the laboratory. Reactions with carbon nucleophiles will be dealt with in chapters 13 and 14, however - in this chapter and the next, we will concentrate on non-carbon nucleophiles.

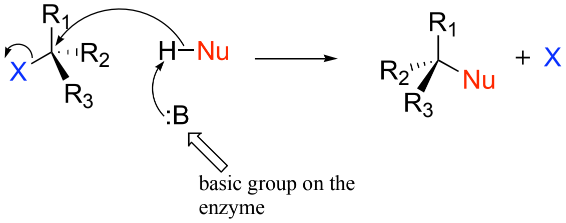
When thinking about nucleophiles, the first thing to recognize is that, for the most part, the same quality of 'electron-richness' that makes a something nucleophilic also makes it basic: *nucleophiles can be bases, and bases can be nucleophiles*. It should not be surprising, then, that most of the trends in basicity that we have already discussed also apply to nucleophilicity.

Now, lets discuss some of the major factors that affect nucleophilicity.

8.3B: Protonation states and nucleophilicity

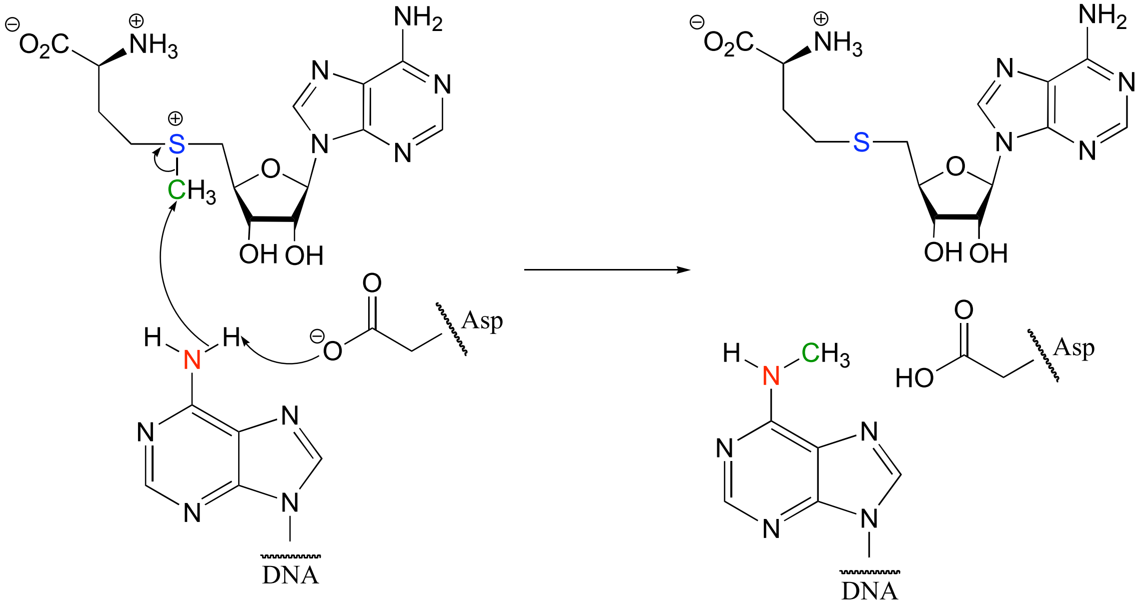
The protonation state of a nucleophilic atom has a very large effect on its nucleophilicity. This is an idea that makes intuitive sense: a hydroxide ion is much more nucleophilic (and basic) than a water molecule, because the negatively charged oxygen on the hydroxide ion carries greater electron density than the oxygen atom of a neutral water molecule. In practical terms, this means that a hydroxide nucleophile will react in an S_N2 reaction with methyl bromide much faster (about 10,000 times faster) than a water nucleophile.

A neutral amine is nucleophilic, whereas a protonated ammonium cation is not. This is why enzymes which have evolved to catalyze nucleophilic reactions often have a basic amino acid side chain poised in position to accept a proton from the nucleophilic atom as the nucleophilic attack occurs.



Depending on the specific reaction being discussed, deprotonation of the nucleophile might occur before, during, or after the actual nucleophilic attack. In general for enzymatic reactions, however, it is most accurate to depict the proton abstraction and nucleophilic attack occurring simultaneously.

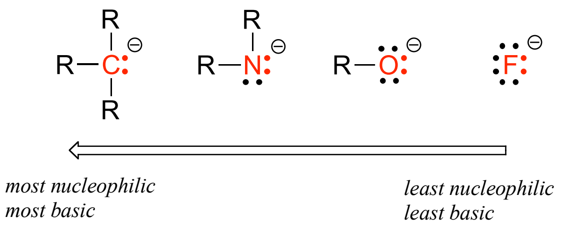
The basic enzymatic group could be a histidine, a neutral (deprotonated) arginine or lysine, or a negatively-charged (deprotonated) aspartate, glutamate, or tyrosine. For example, a more complete picture of the DNA methylation reaction we saw in section 8.1 shows an aspartate from the enzyme's active site accepting a proton from the nucleophilic amine as it attacks the carbon electrophile.



As it is deprotonated by the aspartate, the amine nitrogen becomes more electron-rich, and therefore more nucleophilic.

8.3C: Periodic trends and solvent effects in nucleophilicity

There are predictable periodic trends in nucleophilicity. Moving horizontally across the second row of the table, the trend in nucleophilicity parallels the trend in basicity:

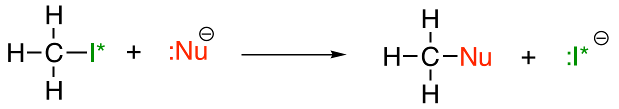


The reasoning behind the horizontal nucleophilicity trend is the same as the reasoning behind the basicity trend: more electronegative elements hold their electrons more tightly, and are less able to donate them to form a new bond.

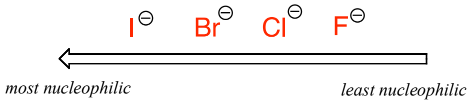
This horizontal trends also tells us that amines are more nucleophilic than alcohols, although both groups commonly act as nucleophiles in both laboratory and biochemical reactions.

Recall from section 7.3A that the basicity of atoms decreases as we move vertically down a column on the periodic table: thiolate ions are less basic than alkoxide ions, for example, and bromide ion is less basic than chloride ion, which in turn is less basic than fluoride ion. Recall also that this trend can be explained by considering the increasing size of the 'electron cloud' around the larger ions: the electron density inherent in the negative charge is spread around a larger area, which tends to increase stability (and thus reduce basicity).

The vertical periodic trend for nucleophilicity is somewhat more complicated that that for basicity: depending on the solvent that the reaction is taking place in, the nucleophilicity trend can go in either direction. Let's take the simple example of the S_N2 reaction below:



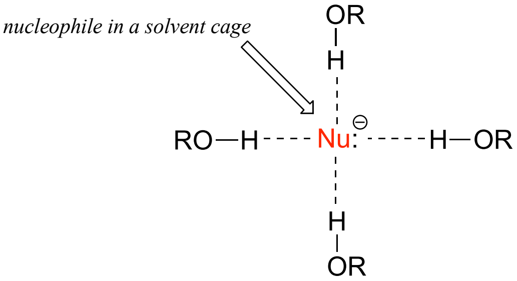
. . .where Nu^- is one of the halide ions: fluoride, chloride, bromide, or iodide, and the leaving group I* is a radioactive isotope of iodine (which allows us to distinguish the leaving group from the nucleophile in that case where both are iodide). If this reaction is occurring in a **protic solvent** (that is, a solvent that has a hydrogen bonded to an oxygen or nitrogen - water, methanol and ethanol are the most important examples), then the reaction will go fastest when iodide is the nucleophile, and slowest when fluoride is the nucleophile, reflecting the relative strength of the nucleophile.



Relative nucleophilicity in protic solvents

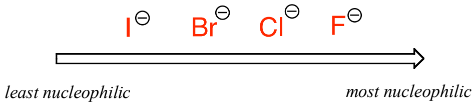
This of course, is opposite that of the vertical periodic trend for basicity, where iodide is the *least* basic (you may want to review the reasoning for this trend in section 7.3A). What is going on here? Shouldn't the stronger base, with its more reactive unbonded valence electrons, also be the stronger nucleophile?

As mentioned above, it all has to do with the solvent. Remember, we are talking now about the reaction running in a *protic* solvent like ethanol. Protic solvent molecules form very strong ion-dipole interactions with the negatively-charged nucleophile, essentially creating a 'solvent cage' around the nucleophile:



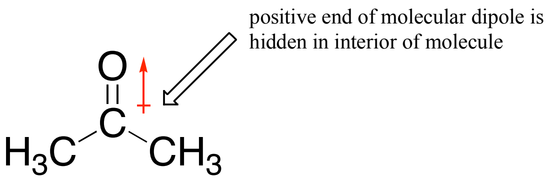
In order for the nucleophile to attack the electrophile, it must break free, at least in part, from its solvent cage. The lone pair electrons on the larger, less basic iodide ion interact less tightly with the protons on the protic solvent molecules - thus the iodide nucleophile is better able to break free from its solvent cage compared the smaller, more basic fluoride ion, whose lone pair electrons are bound more tightly to the protons of the cage.

The picture changes if we switch to a **polar aprotic solvent**, such as acetone, in which there is a molecular dipole but *no hydrogens bound to oxygen or nitrogen*. Now, fluoride is the best nucleophile, and iodide the weakest.



Relative nucleophilicity in polar aprotic solvents

The reason for the reversal is that, with an aprotic solvent, the ion-dipole interactions between solvent and nucleophile are much weaker: the positive end of the solvent's dipole is hidden in the interior of the molecule, and thus it is shielded from the negative charge of the nucleophile.

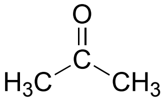


positive end of molecular dipole is hidden in interior of molecule

acetone

A weaker solvent-nucleophile interaction means a weaker solvent cage for the nucleophile to break through, so the solvent effect is much less important, and the more basic fluoride ion is also the better nucleophile.

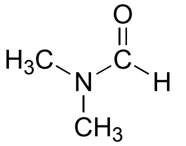
Why not use a completely nonpolar solvent, such as hexane, for this reaction, so that the solvent cage is eliminated completely? The answer to this is simple - the nucleophile needs to be in solution in order to react at an appreciable rate with the electrophile, and a solvent such as hexane will not solvate an a charged (or highly polar) nucleophile at all. That is why chemists use polar aprotic solvents for nucleophilic substitution reactions in the laboratory: they are polar enough to solvate the nucleophile, but not so polar as to lock it away in an impenetrable solvent cage. In addition to acetone, three other commonly used polar aprotic solvents are acetonitrile, dimethylformamide (DMF), and dimethyl sulfoxide (DMSO).



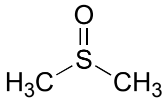
acetone

$N\equiv C-CH_3$

acetonitrile



dimethylformamide (DMF)

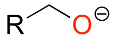


dimethylsulfoxide (DMSO)

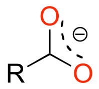
In biological chemistry, where the solvent is protic (water), the most important implication of the periodic trends in nucleophilicity is that thiols are more powerful nucleophiles than alcohols. The thiol group in a cysteine amino acid, for example, is a powerful nucleophile and often acts as a nucleophile in enzymatic reactions, and of course negatively-charged thiolates (RS^-) are even more nucleophilic. This is not to say that the hydroxyl groups on serine, threonine, and tyrosine do not also act as nucleophiles - they do.

8.3D: Resonance effects on nucleophilicity

Resonance effects also come into play when comparing the inherent nucleophilicity of different molecules. The reasoning involved is the same as that which we used to understand resonance effects on basicity (see section 7.3C). If the electron lone pair on a heteroatom is delocalized by resonance, it is inherently less reactive - meaning less nucleophilic, and also less basic. An alkoxide ion, for example, is more nucleophilic and more basic than a carboxylate group, even though in both cases the nucleophilic atom is a negatively charged oxygen. In the alkoxide, the negative charge is localized on a single oxygen, while in the carboxylate the charge is delocalized over two oxygen atoms by resonance.

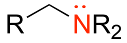


alkoxide ion:
charge is localized
more nucleophilic

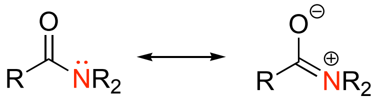


carboxylate ion:
charge is spread over both oxygens
less nucleophilic

The nitrogen atom on an amide is less nucleophilic than the nitrogen of an amine, due to the resonance stabilization of the nitrogen lone pair provided by the amide carbonyl group.



amine
more nucleophilic



amide
less nucleophilic

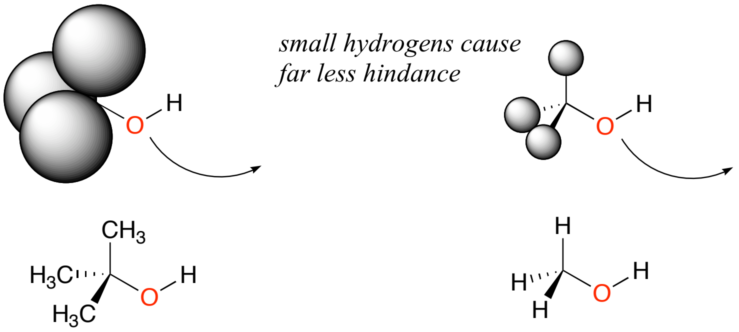
Exercise 8.7: Which amino acid has the more nucleophilic side chain - serine or tyrosine? Explain.

8.3E: Steric effects on nucleophilicity

Steric hindrance is an important consideration when evaluating nucleophility. For example, tert-butanol is less potent as a nucleophile than methanol. This is because the comparatively bulky methyl groups on the tertiary alcohol effectively block the route of attack by the nucleophilic oxygen, slowing the reaction down considerably (imagine trying to walk through a narrow doorway while carrying three large suitcases!).

large methyl groups hinder nucleophlic attack

small hydrogens cause far less hindance



It is not surprising that it is more common to observe serines acting as nucleophiles in enzymatic reactions compared to threonines - the former is a primary alcohol, while the latter is a secondary alcohol.

Exercise 8.8: Which is the better nucleophile - a cysteine side chain or a methionine side chain? Explain.

Exercise 8.9: In each of the following pairs of molecules/ions, which is the better nucleophile in a reaction with CH_3Br in acetone solvent? Explain your choice.

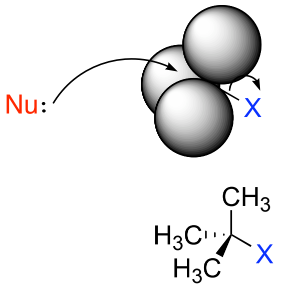
a) phenolate ion (deprotonated phenol) or benzoate ion (deprotonated benzoic acid)
b) water and hydronium ion
c) trimethylamine and triethylamine
d) chloride anion and iodide anion
e) CH_3NH^- and $CH_3CH_2NH_2$

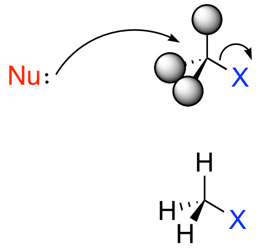
Section 8.4: Electrophiles and carbocation stability

In the vast majority of the nucleophilic substitution reactions you will see in this and other organic chemistry texts, the electrophilic atom is a carbon which is bonded to an electronegative atom, usually oxygen, nitrogen, sulfur, or a halogen. The concept of electrophilicity is relatively simple: an electron-poor atom is an attractive target for something that is electron-rich, *i.e.* a nucleophile. However, we must also consider the effect of steric hindrance on electrophilicity. In addition, we must discuss how the nature of the electrophilic carbon, and more specifically the stability of a potential carbocationic intermediate, influences the S_N1 vs. S_N2 character of a nucleophilic substitution reaction.

8.4A: Steric effects on electrophilicity

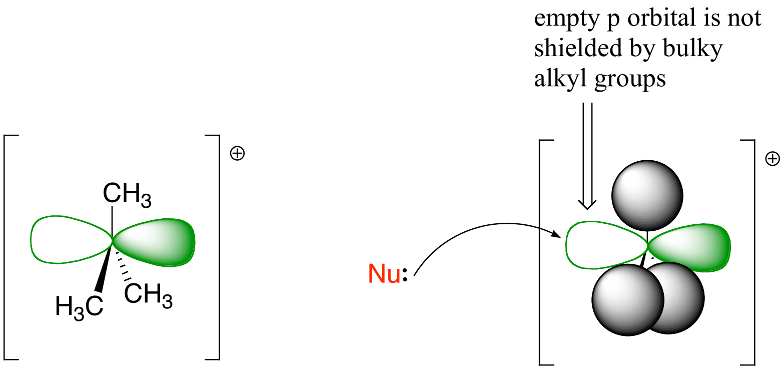
Consider two hypothetical S_N2 reactions: one in which the electrophile is a methyl carbon and another in which it is tertiary carbon.





Because the three substituents on the methyl carbon electrophile are tiny hydrogens, the nucleophile has a relatively clear path for backside attack. However, backside attack on the tertiary carbon is blocked by the bulkier methyl groups. Once again, steric hindrance - this time caused by bulky groups attached to the electrophile rather than to the nucleophile - hinders the progress of an associative nucleophilic (S_N2) displacement.

The factors discussed in the above paragraph, however, do not prevent a sterically-hindered carbon from being a good electrophile - they only make it less likely to be attacked in a *concerted S_N2 reaction*. Nucleophilic substitution reactions in which the electrophilic carbon is sterically hindered are more likely to occur by a two-step, dissociative (S_N1) mechanism. This makes perfect sense from a geometric point of view: the limitations imposed by sterics are significant mainly in an S_N2 displacement, when the electrophile being attacked is a sp^3-hybridized tetrahedral carbon with its relatively 'tight' angles of 109.4°. Remember that in an S_N1 mechanism, the nucleophile attacks an sp^2-hybridized carbocation intermediate, which has trigonal planar geometry with 'open' 120 angles.



With this open geometry, the empty p orbital of the electrophilic carbocation is no longer significantly shielded from the approaching nucleophile by the bulky alkyl groups. A carbocation is a very potent electrophile, and the nucleophilic step occurs very rapidly compared to the first (ionization) step.

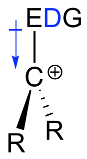
8.4B: Stability of carbocation intermediates

We know that the rate-limiting step of an S_N1 reaction is the first step - formation of the this carbocation intermediate. The rate of this step – and therefore, the rate of the overall substitution reaction – depends on the activation energy for the process in which the bond between the carbon and the leaving group breaks and a carbocation forms. According to Hammond's postulate (section 6.2B), the more stable the carbocation intermediate is, the faster this first bond-breaking step will occur. In other words, the likelihood of a nucleophilic substitution reaction proceeding by a dissociative (S_N1) mechanism depends to a large degree on the stability of the carbocation intermediate that forms.

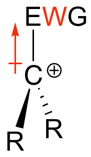
The critical question now becomes, *what stabilizes a carbocation*?

Think back to Chapter 7, when we were learning how to evaluate the strength of an acid. The critical question was "how stable is the conjugate base that results when this acid donates its proton"? In many cases, this conjugate base was an anion – a center of excess electron density. Anything that can draw some of this electron density away– in other words, any electron withdrawing group – will stabilize the anion.

So if it takes an electron *withdrawing* group to stabilize a negative charge, what will stabilize a positive charge? An electron *donating* group!



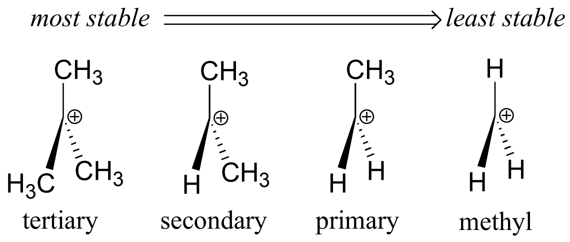
Electron Donating Group:
stabilizes a carbocation



Electron Withdrawing Group:
*de*stabilizes a carbocation

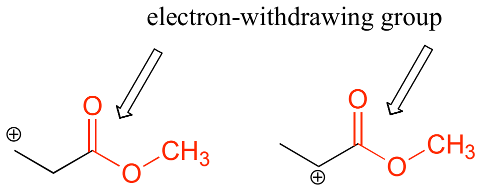
A positively charged species such as a carbocation is very electron-poor, and thus anything which donates electron density to the center of electron poverty will help to stabilize it. Conversely, a carbocation will be *destabilized* by an electron withdrawing group.

Alkyl groups – methyl, ethyl, and the like – are weak electron donating groups, and thus stabilize nearby carbocations. What this means is that, in general, *more substituted carbocations are more stable*: a tert-butyl carbocation, for example, is more stable than an isopropyl carbocation. Primary carbocations are highly unstable and not often observed as reaction intermediates; methyl carbocations are even less stable.



Alkyl groups are electron donating and carbocation-stabilizing because the electrons around the neighboring carbons are drawn towards the nearby positive charge, thus slightly reducing the electron poverty of the positively-charged carbon.

It is not accurate to say, however, that carbocations with higher substitution are *always* more stable than those with less substitution. Just as electron-donating groups can stabilize a carbocation, electron-withdrawing groups act to destabilize carbocations. Carbonyl groups are electron-withdrawing by inductive effects, due to the polarity of the C=O double bond. It is possible to demonstrate in the laboratory (see section 16.1D) that carbocation A below is more stable than carbocation B, even though A is a primary carbocation and B is secondary.



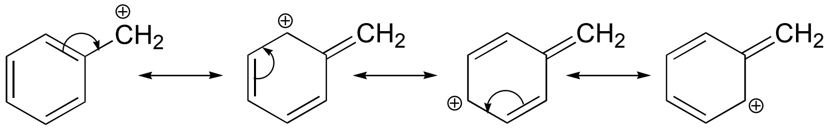
A: more stable

B: less stable

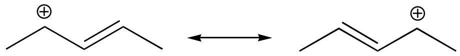
The difference in stability can be explained by considering the electron-withdrawing inductive effect of the ester carbonyl. Recall that inductive effects - whether electron-withdrawing or donating - are relayed through covalent bonds and that the strength of the effect decreases rapidly as the number of intermediary bonds increases. In other words, the effect decreases with distance. In species B the positive charge is closer to the carbonyl group, thus the destabilizing electron-withdrawing effect is stronger than it is in species A.

In the next chapter we will see how the carbocation-destabilizing effect of electron-withdrawing fluorine substituents can be used in experiments designed to address the question of whether a biochemical nucleophilic substitution reaction is S_N1 or S_N2.

Stabilization of a carbocation can also occur through resonance effects, and as we have already discussed in the acid-base chapter, resonance effects as a rule are more powerful than inductive effects. Consider the simple case of a **benzylic** carbocation:

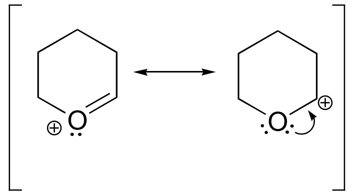


This carbocation is comparatively stable. In this case, electron donation is a resonance effect. Three additional resonance structures can be drawn for this carbocation in which the positive charge is located on one of three aromatic carbons. The positive charge is not isolated on the benzylic carbon, rather it is delocalized around the aromatic structure: this delocalization of charge results in significant stabilization. As a result, benzylic and **allylic** carbocations (where the positively charged carbon is conjugated to one or more non-aromatic double bonds) are significantly more stable than even tertiary alkyl carbocations.

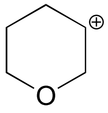


an allylic carbocation

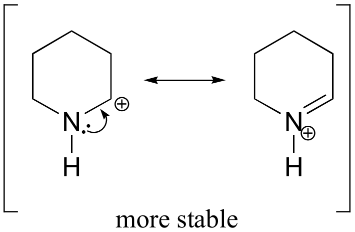
Because heteroatoms such as oxygen and nitrogen are more electronegative than carbon, you might expect that they would by definition be electron withdrawing groups that destabilize carbocations. In fact, the opposite is often true: if the oxygen or nitrogen atom is in the correct position, the overall effect is carbocation stabilization. This is due to the fact that although these heteroatoms are electron *withdrawing* groups by induction, they are electron *donating* groups by resonance, and it is this resonance effect which is more powerful. (We previously encountered this same idea when considering the relative acidity and basicity of phenols and aromatic amines in section 7.4). Consider the two pairs of carbocation species below:



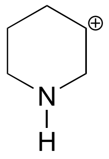
more stable



less stable
(no resonance delocalization)



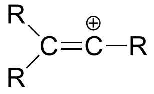
more stable



less stable
(no resonance delocalization)

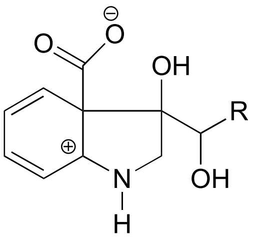
In the more stable carbocations, the heteroatom acts as an electron donating group by resonance: in effect, the lone pair on the heteroatom is available to delocalize the positive charge. In the less stable carbocations the positively-charged carbon is more than one bond away from the heteroatom, and thus no resonance effects are possible. In fact, in these carbocation species the heteroatoms actually *destabilize* the positive charge, because they are electron withdrawing by induction.

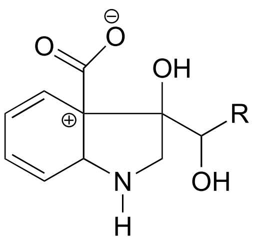
Finally, **vinylic** carbocations, in which the positive charge resides on a double-bonded carbon, are very unstable and thus unlikely to form as intermediates in any reaction.



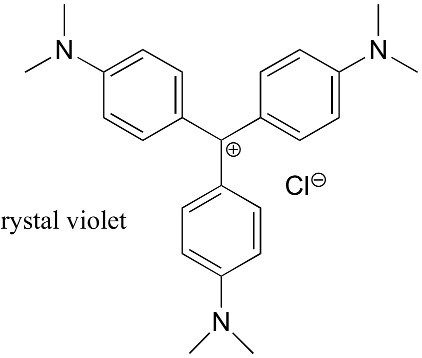
a vinylic carbocation
(very unstable)

Exercise 8.10: In which of the structures below is the carbocation expected to be more stable? Explain.





For the most part, carbocations are very high-energy, transient intermediate species in organic reactions. However, there are some unusual examples of very stable carbocations that take the form of organic salts. Crystal violet is the common name for the chloride salt of the carbocation whose structure is shown below. Notice the structural possibilities for extensive resonance delocalization of the positive charge, and the presence of three electron-donating amine groups.

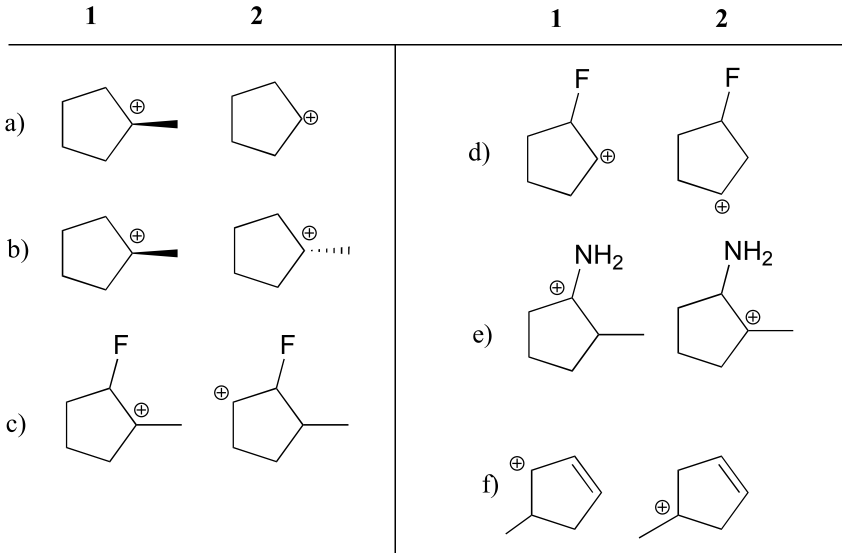


crystal violet

Exercise 8.11: Draw a resonance structure of the crystal violet cation in which the positive charge is delocalized to one of the nitrogen atoms.

When considering the possibility that a nucleophilic substitution reaction proceeds *via* an S_N1 pathway, it is critical to evaluate the stability of the hypothetical carbocation intermediate. If this intermediate is not sufficiently stable, an S_N1 mechanism must be considered unlikely, and the reaction probably proceeds by an S_N2 mechanism. In the next chapter we will see several examples of biologically important S_N1 reactions in which the positively charged intermediate is stabilized by inductive and resonance effects inherent in its own molecular structure.

Exercise 8.12: State which carbocation in each pair below is more stable, or if they are expected to be approximately equal. Explain your reasoning.



Section 8.5: Leaving groups

8.5A: What makes a good leaving group?

In our general discussion of nucleophilic substitution reactions, we have until now been designating the leaving group simply as "X". As you may imagine, however, the nature of the leaving group is an important consideration: if the C-X bond does not break, the new bond between the nucleophile and electrophilic carbon cannot form, regardless of whether the substitution is S_N1 or S_N2.

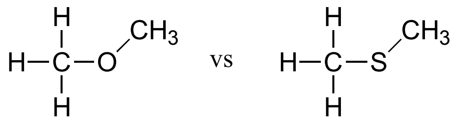
When the C-X bond breaks in a nucleophilic substitution, the pair of electrons in the bond goes with the leaving group. In this way, the leaving group is analogous to the conjugate base in a Brønsted-Lowry acid-base reaction. When we were evaluating the strength of acids in chapter 7, what we were really doing was evaluating the stability of the conjugate base that resulted from the proton transfer. All of the concepts that we used to evaluate the stability of conjugate bases we can use again to evaluate leaving groups – essentially they are one and the same. In other words, the trends in basicity are parallel to the trends in leaving group potential - the weaker the base, the better the leaving group. Just as with conjugate bases, the most important question regarding leaving groups is this: when a leaving group leaves and takes a pair of electrons with it, how well is the extra electron density stabilized?

In laboratory synthesis reactions, halides often act as leaving groups. Iodide, which is the *least* basic of the four main halides, is also the *best* leaving group – it is the most stable as a negative ion. Fluoride is the least effective leaving group among the halides, because fluoride anion is the most basic.

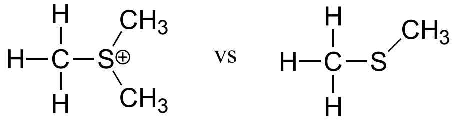
Exercise 8.13: In each pair (A and B) below, which electrophile would be expected to react more rapidly in an S_N2 reaction with a given nucleophile?

A **B** **A** **B**

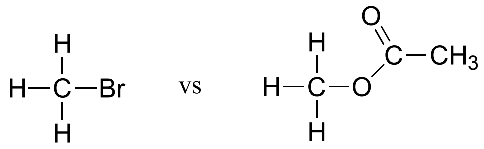
a)



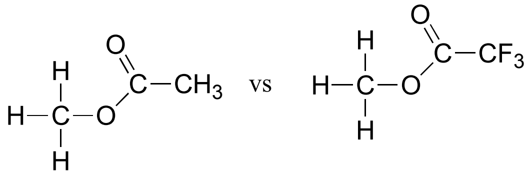
b)



c)



d)

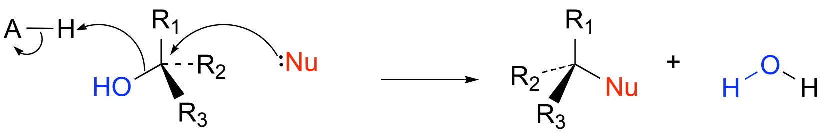


We already know (section 8.3C) that the use of polar, aprotic solvents increases the reactivity of nucleophiles in S_N2 reactions, because these solvents do not 'cage' the nucleophile and keep it from attacking the electrophile.

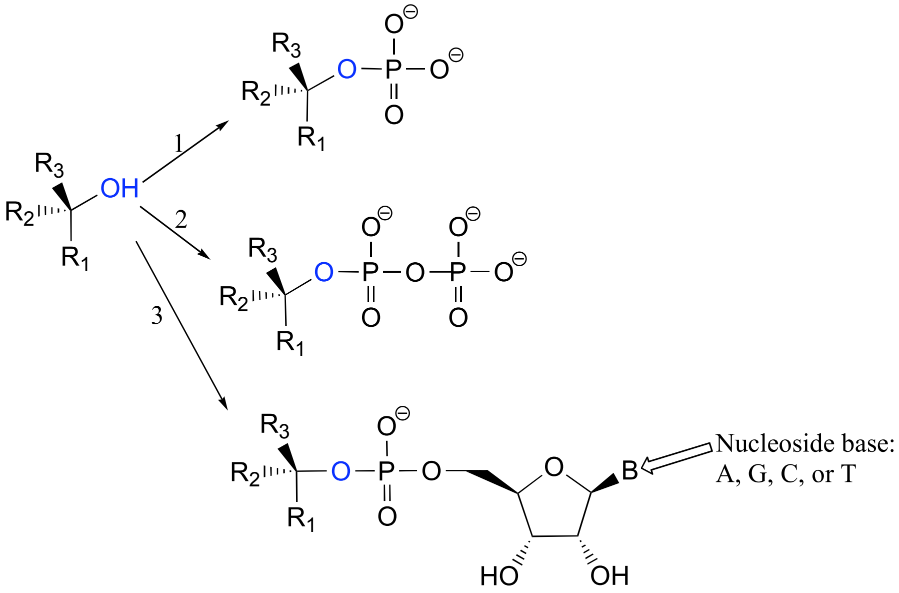
On the other hand, the rates of S_N1 reactions are generally increased by the use of a highly polar solvent, including protic (hydrogen bonding) solvents such as water or ethanol. In essence, a protic solvent increases the reactivity of the leaving group in an S_N1 reaction, by helping to stabilize the products of the first (ionization) step. In the S_N1 mechanism, remember, the rate determining step does *not* involve the nucleophilic species, so any reduction of nucleophilicity does not matter. What matters is that the charged products of the first step - the carbocation intermediate and the anionic leaving group - are stabilized best by a highly polar, protic solvent.

8.5B: Leaving groups in biochemical reactions

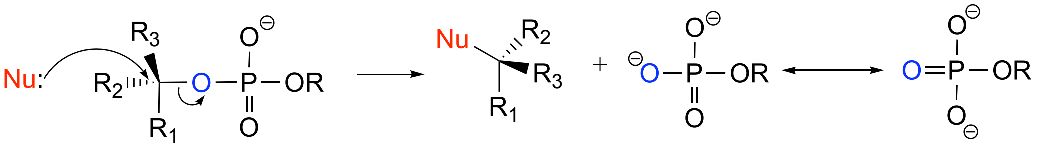
In biological reactions, we do not often see halides serving as leaving groups (in fact, outside of some marine organisms, halogens are fairly unusual in biological molecules). More common leaving groups in biochemical reactions are phosphates, water, alcohols, and thiols. In many cases, the leaving group is protonated by an acidic group on the enzyme as bond-breaking occurs. For example, hydroxide ion itself seldom acts as a leaving group – it is simply too high in energy (too basic). Rather, the hydroxide oxygen is generally protonated by an enzymatic acid before or during the bond-breaking event, resulting in a (very stable) water leaving group.



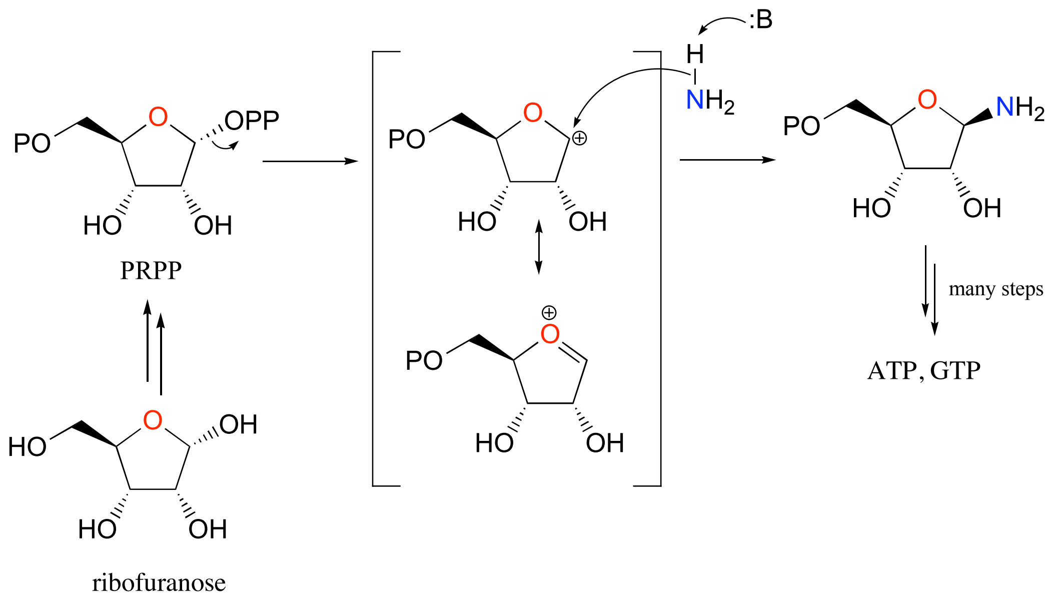
More often, however, the hydroxyl group of an alcohol is first converted enzymatically to a phosphate ester in order to create a better leaving group. This phosphate ester can take the form of a simple monophosphate (arrow 1 in the figure below), a diphosphate (arrow 2), or a nucleotide monophosphate (arrow 3).



Due to resonance delocalization of the developing negative charge, phosphates are excellent leaving groups.



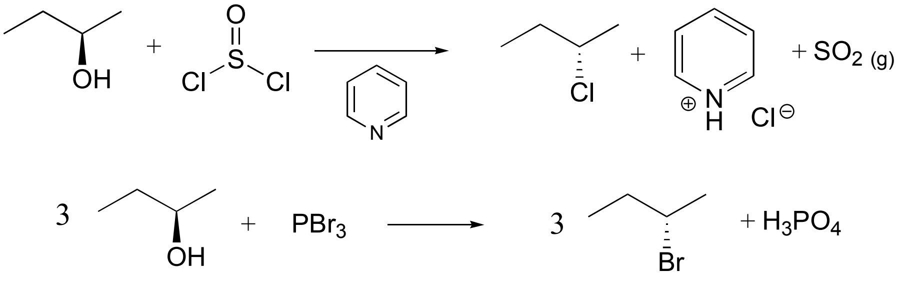
Here's a specific example (from DNA nucleotide biosynthesis) that we will encounter in more detail in section 11.5:



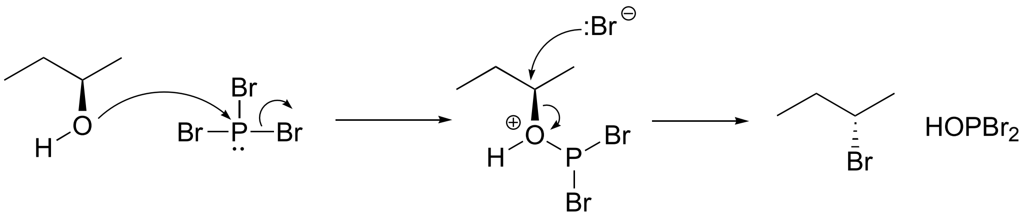
Here, the OH group on ribofuranose is converted to a diphosphate, a much better leaving group. Ammonia is the nucleophile in the second step of this S_N1-like reaction.
We will learn much more about phosphates in chapter 10. What is important for now is that in each case, an alcohol has been converted into a much better leaving group, and is now primed for a nucleophilic substitution reaction.

8.5C: Synthetic parallel - conversion of alcohols to alkyl halides, tosylates and mesylates

Synthetic organic chemists, when they want to convert an alcohol into a better leaving group, have several methods to choose from. One common strategy is to convert the alcohol into an alkyl chloride or bromide, using thionyl chloride or phosphorus tribromide:

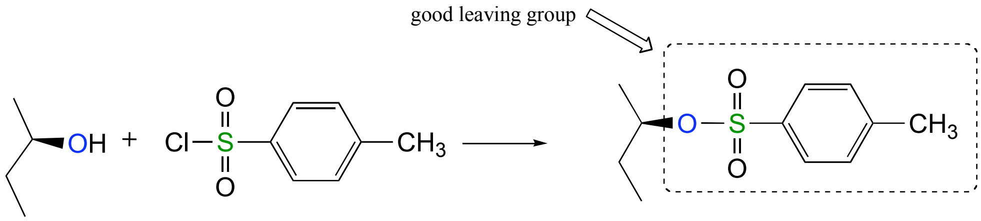


We won’t worry yet about the mechanism for the thionyl chloride reaction (you’ll have a chance to propose a mechanism for this reaction in the chapter 12 problems), but the PBr_3 reaction is thought to involve two successive S_N2-like steps:

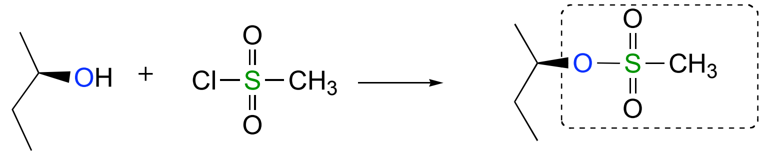


Notice that these reactions result in inversion of stereochemistry in the resulting alkyl halide.

Alternatively, we can transform an alcohol group into sulfonic ester using *para*-toluene sulfonyl chloride (Ts-Cl) or methanesulfonyl chloride (Ms-Cl), creating what is termed an organic **tosylate** or **mesylate:**



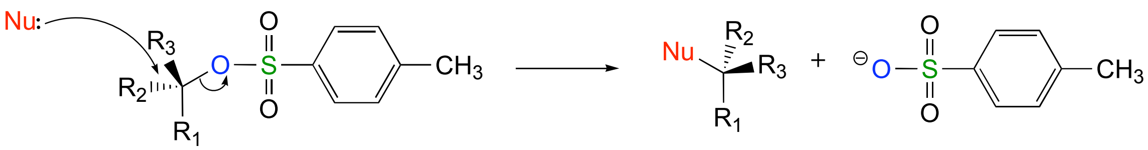
an organic tosylate (R-OTs)



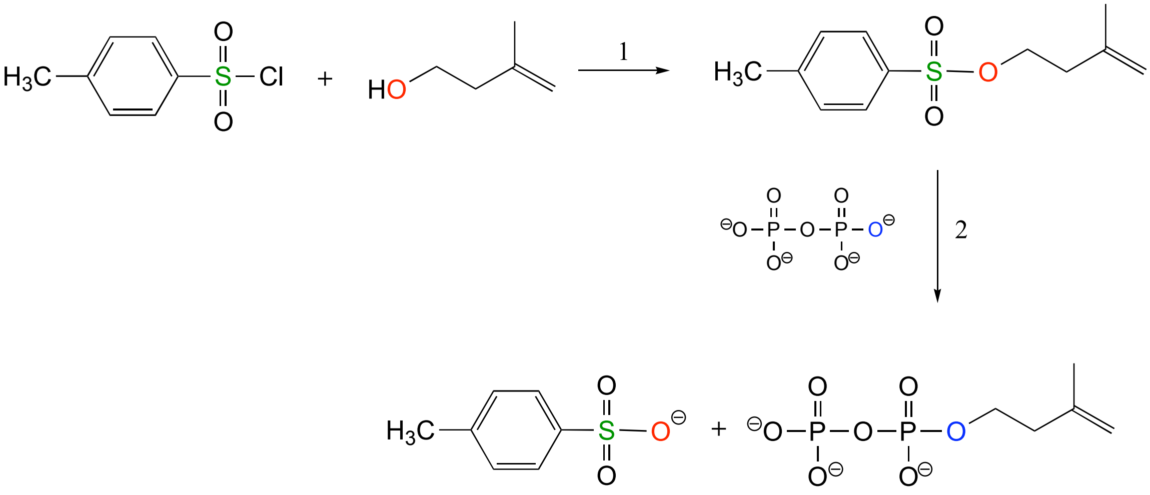
an organic mesylate (R-OMs)

Again, you'll have a chance to work a mechanism for tosylate and mesylate formation in the chapter 12 problems. Notice, though, that unlike the halogenation reactions above, conversion of an alcohol to a tosylate or mesylate proceeds with retention of configuration at the electrophilic carbon.

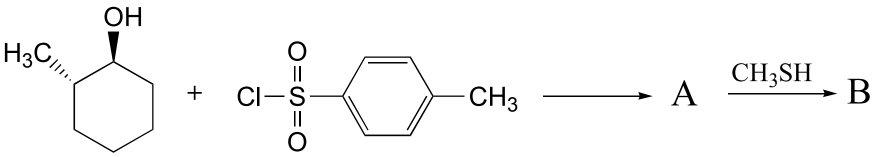
Chlorides, bromides, and tosylate / mesylate groups are excellent leaving groups in nucleophilic substitution reactions, due to resonance delocalization of the developing negative charge on the leaving oxygen.



The laboratory synthesis of isopentenyl diphosphate - the 'building block' molecule used by nature for the construction of isoprenoid molecules such as cholesterol and β-carotene - was accomplished by first converting the alcohol into an organic tosylate (step 1), then displacing the tosylate group with an inorganic pyrophosphate nucleophile (step 2) (*J. Org. Chem* **1986**, *51*, 4768).



Exercise 8.14: Predict the structures of A and B in the following reaction:



8.5D: S_N1 or S_N2? Predicting the mechanism

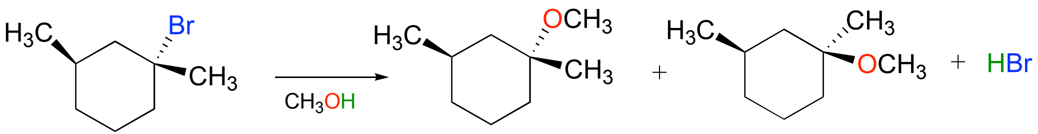
When considering whether a nucleophilic substitution is likely to occur via an S_N1 or S_N2 mechanism, we really need to consider three factors:

1) **The electrophile:** when the leaving group is attached to a methyl group or a primary carbon, an S_N2 mechanism is favored (here the electrophile is unhindered by surrounded groups, and any S_N1 carbocation intermediate would be high-energy and thus unlikely). When the leaving group is attached to a tertiary, allylic, or benzylic carbon, a carbocation intermediate will be relatively stable and thus an S_N1 mechanism is favored.

2) **The nucleophile**: powerful nucleophiles, especially those with negative charges, favor the S_N2 mechanism. Weaker nucleophiles such as water or alcohols favor the S_N1 mechanism.

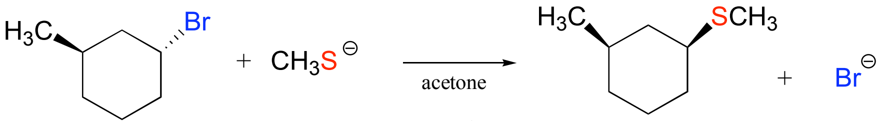
3) **The solvent**: Polar aprotic solvents favor the S_N2 mechanism by enhancing the reactivity of the nucleophile. Polar protic solvents favor the S_N1 mechanism by stabilizing the carbocation intermediate. S_N1 reactions are frequently solvolysis reactions.

For example, the reaction below has a tertiary alkyl bromide as the electrophile, a polar protic solvent (methanol) and a weak nucleophile (also methanol). Thus we'd confidently predict an S_N1 reaction mechanism. Because substitution occurs at a chiral carbon, we can also predict that the reaction will proceed with racemization.



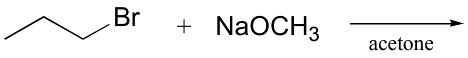
S_N1 mechanism

In the reaction below, on the other hand, the electrophile is a secondary alkyl bromide – in this case, both S_N1 and S_N2 mechanisms are possible, depending on the nucleophile and the solvent. In this example, the nucleophile (a thiolate anion) is strong, and a polar protic solvent is used – so the S_N2 mechanism is heavily favored over S_N1. The reaction is expected to proceed with inversion of configuration.

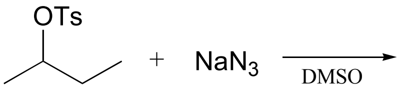


S_N2 mechanism

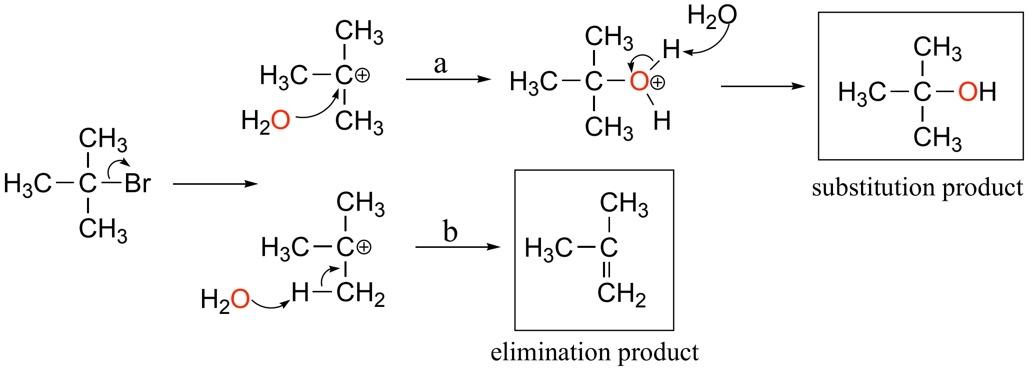
Exercise 8.15: Determine whether each substitution reaction shown below is likely to proceed by an S_N1 or S_N2 mechanism.

a) 

b) $\mathrm{C(CH_3)_3Cl} \xrightarrow[\text{ethanol}]{}$

c) 

In all of our discussion so far about nucleophilic substitutions, we have ignored another important possibility. In many cases, including the two examples above, substitution reactions compete with a type of reaction known as **elimination**. Consider, for example, the two courses that a reaction could take when *tert*-butyl bromide reacts with water:



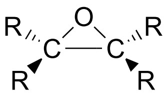
We begin with formation of the carbocation intermediate. In pathway 'a', water acts as a nucleophile – this is, of course, the familiar S_N1 reaction. However, a water molecule encountering the carbocation intermediate could alternatively act as a *base* rather than as a nucleophile, plucking a proton from one of the methyl carbons and causing the formation of a new carbon-carbon π bond. This alternative pathway is called an elimination reaction, and in fact with the conditions above, both the substitution and the elimination pathways will occur in competition with each other.

We will have lots more to say about elimination reactions in chapter 14, focusing on biochemical eliminations but also thinking about the competition between substitution and elimination that occurs with many nonenzymatic reactions.

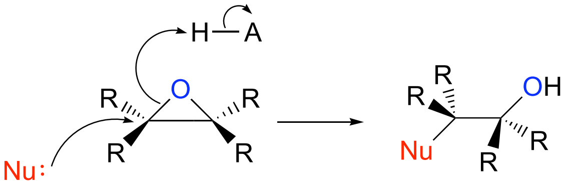
Section 8.6: Epoxides as electrophiles in nucleophilic substitution reactions

8.6A: Epoxide structure

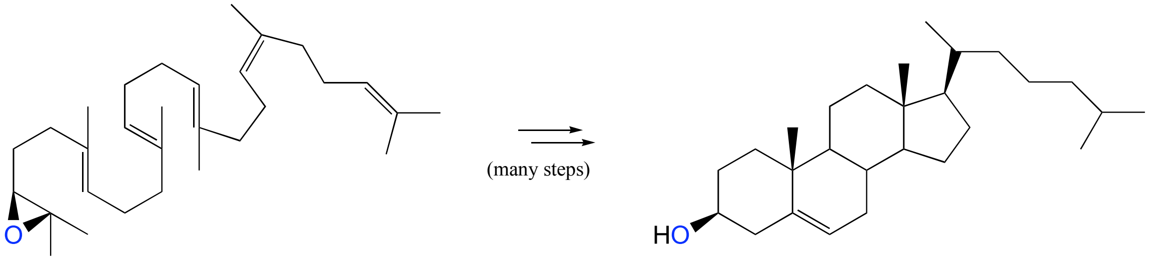
Epoxides (also known as **oxiranes**) are three-membered ring structures in which one of the vertices is an oxygen and the other two are carbons.



The carbons in an epoxide group are very reactive electrophiles, due in large part to the fact that substantial ring strain is relieved when the ring opens upon nucleophilic attack.


three-membered ring: high energy (ring strain)
ring has been opened, energy released

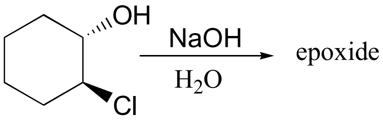
Epoxides are very important intermediates in laboratory organic synthesis, and are also found as intermediate products in some biosynthetic pathways. The compound (3S)-2,3-oxidosqualene, for example, is an important intermediate in the biosynthesis of cholesterol (we'll see the epoxide ring-opening step in chapter 15):



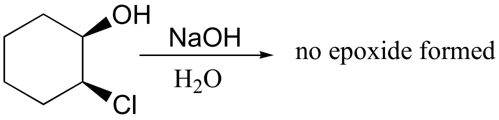
(3S)-2,3-oxidosqualene cholesterol

Both in the laboratory and in the cell, epoxides are usually formed by the oxidation of an alkene. These reactions will be discussed in detail in chapter 16.

Exercise 8.16: Another way to form an epoxide that does not involve an oxidative reaction is to treat a starting material such as compound A below with a base.



A



B

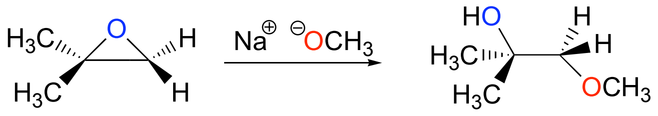
a) Write a mechanism showing the formation of an epoxide from compound A. Show stereochemistry.

b) Compound B does *not* form an epoxide when treated with base. Why do you think this is, and what does this observation tell you about the mechanism of the epoxide-forming reaction of compound A?

8.6B: Epoxide ring-opening reactions - S_N1 vs. S_N2, regioselectivity, and stereoselectivity

The nonenzymatic ring-opening reactions of epoxides provide a nice overview of many of the concepts we have seen already in this chapter. Ring-opening reactions can proceed by either S_N2 or S_N1 mechanisms, depending on the nature of the epoxide and on the reaction conditions. If the epoxide is asymmetric, the structure of the product will vary according to which mechanism dominates. When an asymmetric epoxide undergoes solvolysis in basic methanol, ring-opening occurs by an S_N2 mechanism, and the *less* substituted carbon is the site of nucleophilic attack, leading to what we will refer to as product B:

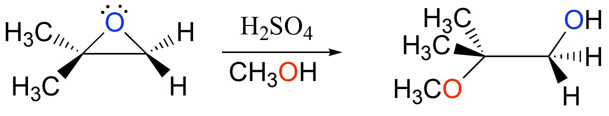
basic ring-opening:



product **B**

Conversely, when solvolysis occurs in acidic methanol, the reaction occurs by a mechanism with substantial S_N1 character, and the *more* substituted carbon is the site of attack. As a result, product A predominates.

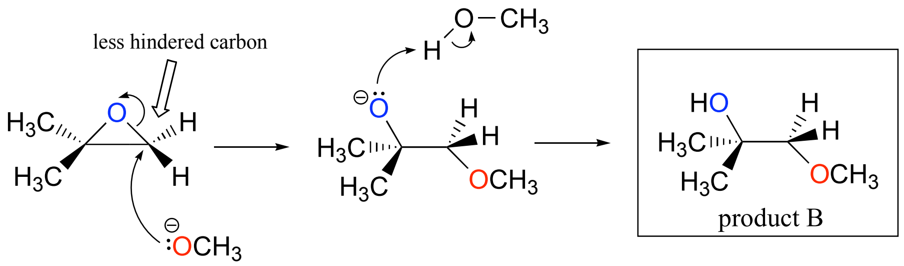
acidic ring-opening:



product **A**

These are both good examples of **regioselective reactions**. In a regioselective reaction, two (or more) different constitutional isomers are possible as products, but one is formed preferentially (or sometimes exclusively).

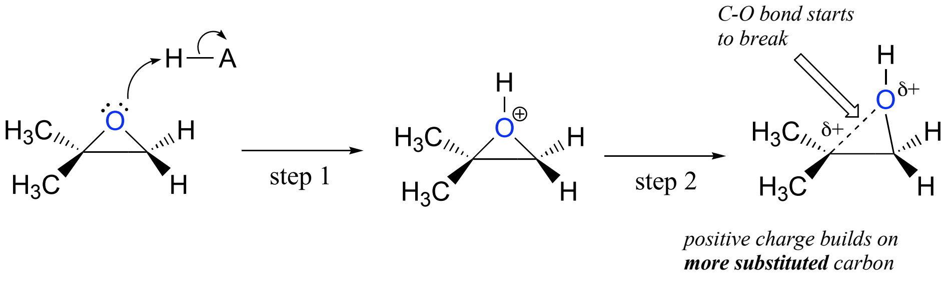
Let us examine the basic, S_N2 case first. The leaving group is an alkoxide anion, because there is no acid available to protonate the oxygen prior to ring opening. An alkoxide is a poor leaving group, and thus the ring is unlikely to open without a 'push' from the nucleophile.



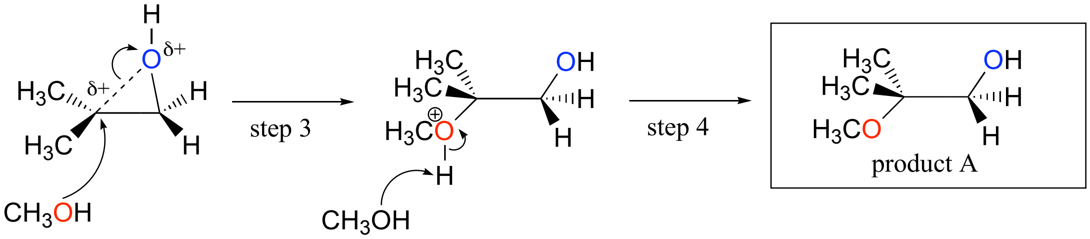
The nucleophile itself is potent: a deprotonated, negatively charged methoxide ion. When a nucleophilic substitution reaction involves a poor leaving group and a powerful nucleophile, it is very likely to proceed by an S_N2 mechanism.

What about the electrophile? There are two electrophilic carbons in the epoxide, but the best target for the nucleophile in an S_N2 reaction is the carbon that is *least hindered*. This accounts for the observed regiochemical outcome. Like in other S_N2 reactions, nucleophilic attack takes place from the backside, resulting in inversion at the electrophilic carbon.

Probably the best way to depict the acid-catalyzed epoxide ring-opening reaction is as a hybrid, or cross, between an S_N2 and S_N1 mechanism. First, the oxygen is protonated, creating a good leaving group (step 1 below) . Then the carbon-oxygen bond begins to break (step 2) and positive charge begins to build up on the more substituted carbon (recall the discussion from section 8.4B about carbocation stability).



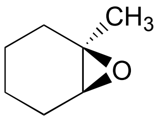
Unlike in an S_N1 reaction, the nucleophile attacks the electrophilic carbon (step 3) before a complete carbocation intermediate has a chance to form.



Attack takes place preferentially from the backside (like in an S_N2 reaction) because the carbon-oxygen bond is still to some degree in place, and the oxygen blocks attack from the front side. Notice, however, how the regiochemical outcome is different from the base-catalyzed reaction: in the acid-catalyzed process, the nucleophile attacks the more substituted carbon because it is this carbon that holds a greater degree of positive charge.

Exercise 8.17: Predict the major product(s) of the ring opening reaction that occurs when the epoxide shown below is treated with:

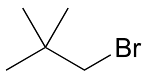
a) ethanol and a small amount of sodium hydroxide
b) ethanol and a small amount of sulfuric acid



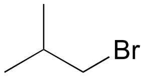
Hint: be sure to consider both regiochemistry ***and*** *stereochemistry!*

Chapter 8 problems

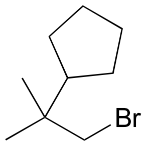
P8.1: Rank the following molecules in order of how fast they would be expected to react with CH_3S Na in acetone:



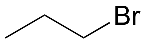
A



B



C



D

P8.2: Draw line structures representing the *most stable* carbocation with the given molecular formula:

a) $C_3H_7^+$

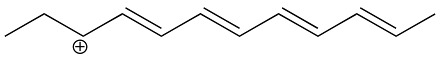
b) $C_4H_9^+$

c) $C_3H_8N^+$

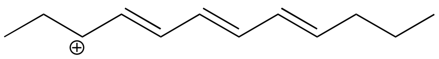
d) $C_4H_7^+$

P8.3: Which of the two carbocations below is expected to be lower in energy? Explain your reasoning.

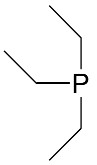
A



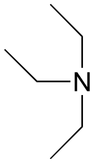
B



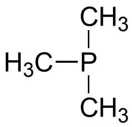
P8.4: Arrange the following species in order of increasing nucleophilicity.



A

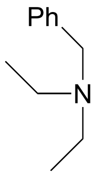


B



$$H_3C-P(CH_3)_2$$

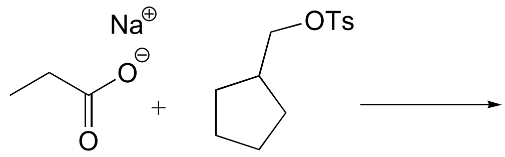
C



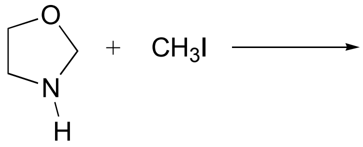
D

P8.5: Predict the organic products of the following S_N2 reactions. Indicate stereochemistry.

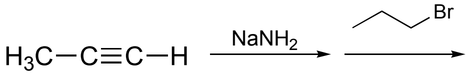
a)



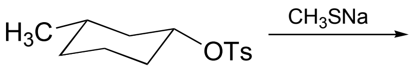
b)



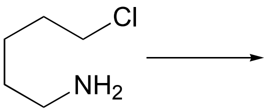
c)



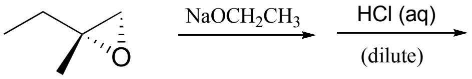
d)



e)



f)



P8.6: Which of the S_N2 reactions in the previous problem is likely to have a *first order* rate expression? Explain.

P8.7: From the following pairs, select the better nucleophile:

a) water or hydroxide ion

b) CH_3S^- or CH_3OH

c) CH_2S^- or CH_3SH

d) acetate ion or hydroxide ion

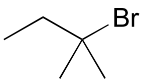
e) diethyl sulfide or diethyl ether

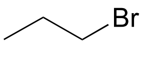
f) dimethylamine or diethylether

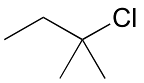
g) trimethylamine or 2,2-dimethylpropane

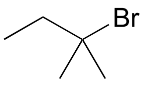
P8.8: Methyl iodide (0.10 mole) is added to a solution that contains 0.10 mole $NaOCH_3$ and 0.10 mole $NaSCH_3$ in ethanol solvent. Predict the major product that would be isolated, and explain your reasoning.

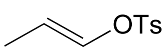
P8.9: For each pair of compounds, predict which will more rapidly undergo S_N1 solvolysis in a water/methanol solvent.

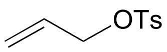
a) 

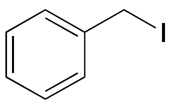
or


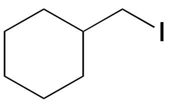
b) 

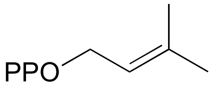
or


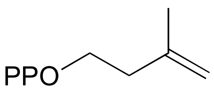
c) 

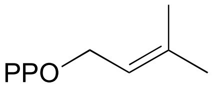
or


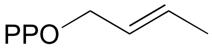
d) 

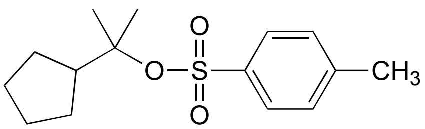
or


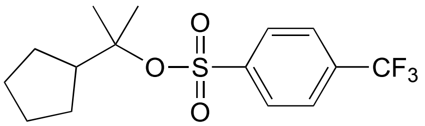
e) 

or


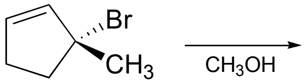
f) 

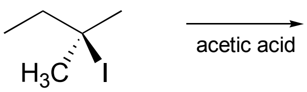
or


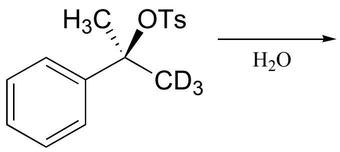
g) 

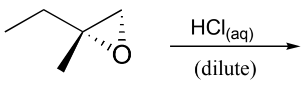
or


P8.10: Predict the solvolysis products (considering stereochemistry) of each of the reactions below. Show the carbocation intermediate in each reaction.

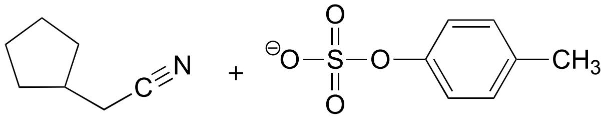
a) 

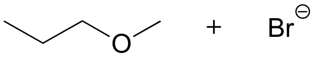
b) 

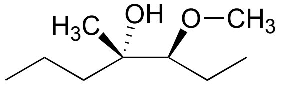
c) 

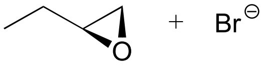
d) 

P8.11: The following compounds are all direct products of nucleophilic substitution reactions. Show the starting compounds.

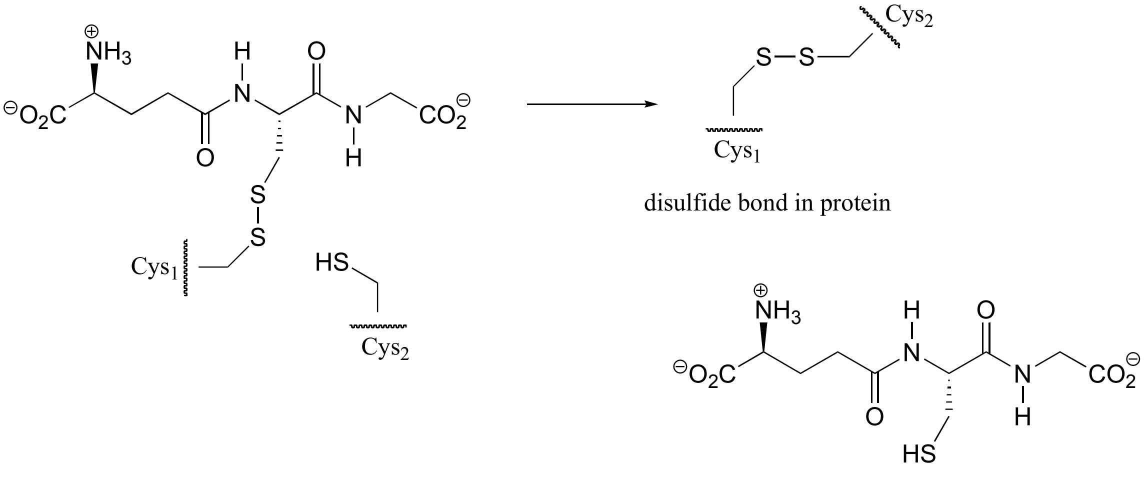
a) 

b) 

c) 

d) 

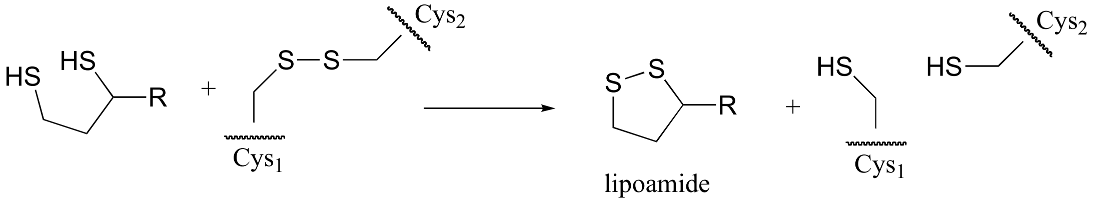
P8.12: The following is the final step in the formation of a 'disulfide bond' between two cysteine residues within a protein (we'll learn more about disulfide bonds in section 16.12A). Propose a mechanism for this process, using H-A and :B to represent catalytic acid and base groups. Label those atoms which are acting as nucleophile, electrophile, and leaving group, respectively.



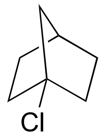
disulfide bond in protein

glutathione

P8.13: The figure below shows the regeneration of the disulfide bond in a coenzyme called lipoamide (we'll learn more about one very important metabolic role of lipoamide in section 16.12B). In the process, a disulfide bond in the enzyme is broken to release two free cysteines. Propose a likely mechanism for this transformation (hint – there are two S_N2 displacements involved).



P8.14: The 'bicyclo' compound shown below is very unreactive in S_N1 reactions, even though it is a tertiary alkyl halide. Explain. (Hint – consider bond geometry - a model will be very helpful!)



Challenge problems

C8.1: In a classic experiment in physical organic chemistry, (*R*)-2-iodooctane was allowed to react with a radioactive isotope of iodide ion, and the researchers monitored how fast the radioactive iodide was incorporated into the alkane (the rate of incorporation, k_i) and also how fast optical activity was lost (the rate of racemization, k_r). They found that the rate of racemization was, within experimental error, equal to twice the rate of incorporation. Discuss the significance of this result - what does it say about the actual mechanism of the reaction?

Chapter 9

Nucleophilic substitution reactions, part II

Introduction

In Chapter 8, you were introduced to the 'nuts and bolts' of the nucleophilic substitution reaction, meeting the 'Three Amigos' of organic reactions (nucleophiles, electrophiles, and leaving groups), learning the difference between S_N1 and S_N2 mechanisms, and thinking about issues of regio- and stereoselectivity. In the interest of keeping these discussions as clear as possible, most examples were very simple and generalized.

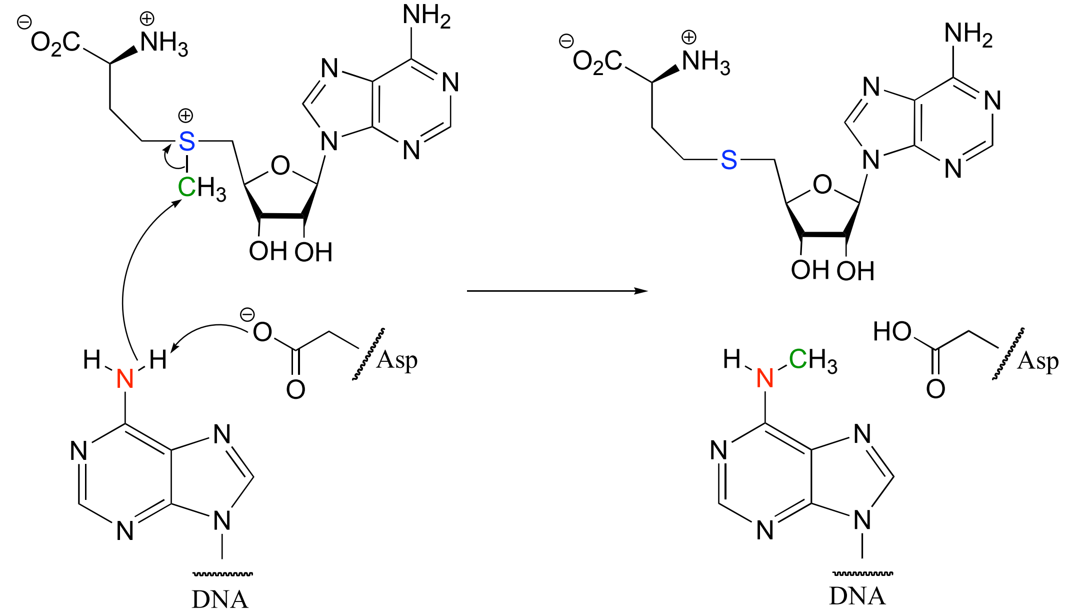
Now that we know the basics, we're going to make things a lot more interesting by looking at real examples of biologically relevant nucleophilic substitutions. We're also going to explore a little bit about *how* some of the ideas from Chapter 8 were discovered: we'll talk about the data from actual experiments performed by scientists who wanted to address questions such as "is this nucleophilic substitution S_N1 or S_N2?", or "can I show that this reaction really proceeds with inversion of configuration?". As we do this, remember always to keep focused on the simple things you already know: for each reaction, be sure that you can strip away in your mind the unreactive parts of the molecules and focus on the interplay between the nucleophile, electrophile, and leaving group.

We'll begin by looking more closely at a biochemical S_N2 reaction that we have already glimpsed briefly in chapters 6 and 8.

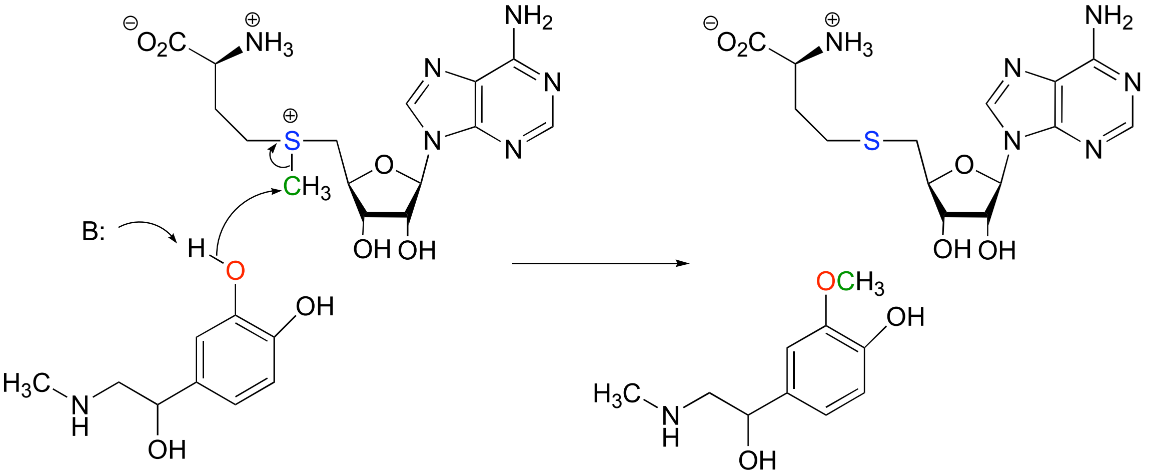
Section 9.1: Methyl group transfers - examples of S_N2 reactions

9.1A: SAM methyltransferases

Some of the most important examples of S_N2 reactions in biochemistry are those catalyzed by S-adenosyl methionine (SAM) – dependent methyltransferase enzymes. We have already seen, in chapter 6 and again in chapter 8, how a methyl group is transferred in an S_N2 reaction from SAM to the amine group on the nucleotide base adenosine:

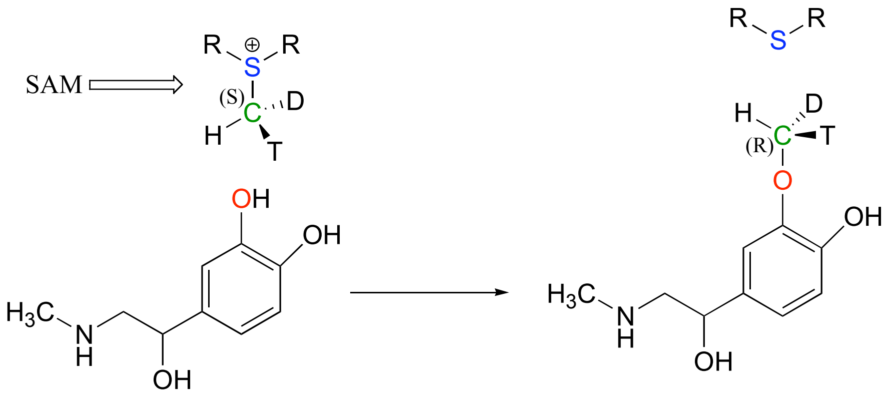


Another SAM-dependent methylation reaction is catalyzed by an enzyme called catechol-O-methyltransferase. The substrate here is epinephrine, also known as adrenaline.



Notice that in this example, the attacking nucleophile is an alcohol rather than an amine (that's why the enzyme is called an O-methyltransferase). In both cases, though, a basic amino acid side chain is positioned in the active site in just the right place to deprotonate the nucleophilic group as it attacks, increasing its nucleophilicity. The electrophile in both reactions is a methyl carbon, so there is little steric hindrance to slow down the nucleophilic attack. The methyl carbon is electrophilic because it is bonded to a positively-charged sulfur, which is a powerful electron withdrawing group. The positive charge on the sulfur also makes it an excellent leaving group, as the resulting product will be a neutral and very stable sulfide. All in all, in both reactions we have a reasonably good nucleophile, an electron-poor, unhindered electrophile, and an excellent leaving group.

Because the electrophilic carbon in these reactions is a methyl carbon, a stepwise S_N1-like mechanism is extremely unlikely: a methyl carbocation is very high in energy and thus is not a reasonable intermediate to propose. We can confidently predict that this reaction is S_N2. Does this S_N2 reaction occur, as expected, with inversion of stereochemistry? Of course, the electrophilic methyl carbon in these reactions is achiral, so inversion is not apparent. To demonstrate inversion, the following experiment has been carried out with catechol-O-methyltransferase:

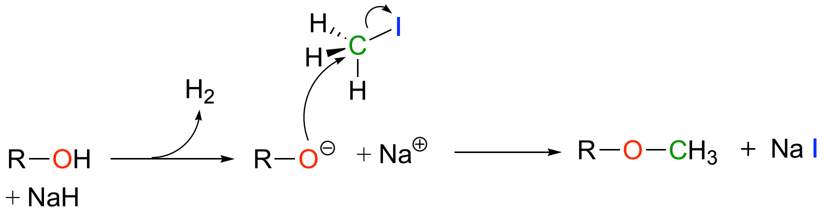


Here, the methyl group of SAM was made to be chiral by incorporating hydrogen isotopes tritium (^{3}H, T) and deuterium (^{2}H, D). The researchers determined that the reaction occurred with inversion of configuration, as expected for an S_N2 displacement (*J. Biol. Chem*. **1980,** *255*, 9124).

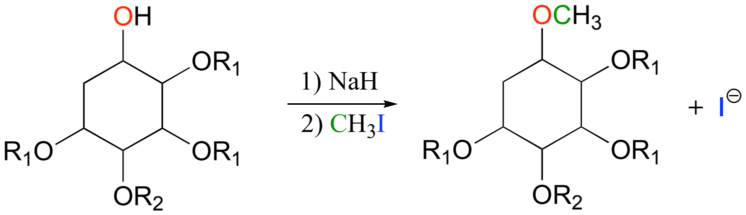
Exercise 9.1: SAM is formed by a nucleophilic substitution reaction between methionine and adenosine triphosphate (ATP). Propose a mechanism for this reaction.

9.1B: Synthetic parallel – the Williamson ether synthesis

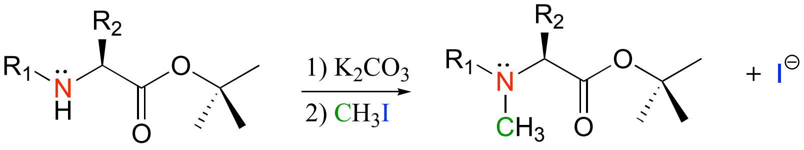
Synthetic organic chemists often use reactions in the laboratory that are conceptually very similar to the SAM-dependent methylation reactions described above. In what is referred to as "O-methylation", an alcohol is first deprotonated by a strong base, typically sodium hydride (this is essentially an irreversible deprotonation, as the hydrogen gas produced can easily be removed from the reaction vessel).



The alkoxide ion, a powerful nucleophile, is then allowed to attack the electrophilic carbon in iodomethane, displacing iodide in an S_N2 reaction. Recall (section 7.3A) that iodide ion is the least basic - and thus the best leaving group - among the halogens commonly used in the synthetic lab. CH_3I is one of the most commonly used lab reagents for methyl transfer reactions, and is the lab equivalent of SAM. In the synthesis of a modified analog of the signaling molecule *myo*-inisitol triphosphate, an alcohol group was methylated using sodium hydride and iodomethane (*Carbohydrate Research* **2004,** *339*, 51):

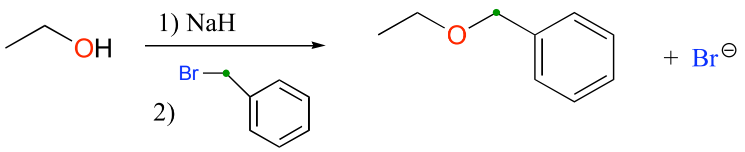


Methylation of amines (N-methylation) by iodomethane is also possible. A recent study concerning the optimization of peptide synthesis methods involved the following reaction (*J. Org. Chem*. **2000** *65*, 2309):



Exercise 9.2: Notice that in the N-methylation reaction a weak base (potassium carbonate) is used, rather than the very strong base (sodium hydride) that was used in the o-methylation reaction. Why is this weak base sufficient for N-methylation but not for o-methylation? In your answer, draw a step-by step mechanism for both reactions, and think carefully about pK_a values and where in each mechanism the deprotonation step occurs.

The methylation of an alcohol by iodomethane is just one example of what is generally referred to as the **Williamson ether synthesis**, in which alcohols are first deprotonated by a strong base and then allowed to attack an alkyl halide electrophile. Below we see the condensation between ethanol and benzyl bromide:

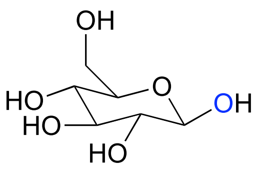


It is important to note that this reaction does not work well at all if secondary or tertiary alkyl halides are used: remember that the alkoxide ion is a strong base as well as a nucleophile, and elimination will compete with nucleophilic substitution (8.5D).

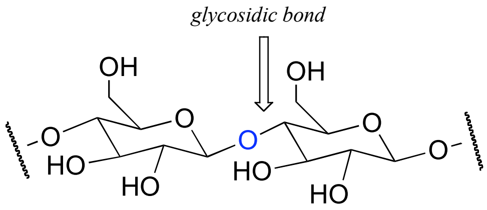
Section 9.2: Digestion of carbohydrate by glycosidase - an S_N1 reaction

A very important example of a nucleophilic substitution reaction that is thought to occur though a two-step, dissociative (S_N1) mechanism is the hydrolytic breakdown of carbohydrates. In chapter 11 we will focus on these reactions and learn that the terms 'hemiacetal', 'acetal', 'hemiketal', and 'ketal' are used to describe the different patterns of bonding that we are seeing here - but for now, we will just look at the reaction from the standpoint of the S_N1 mechanism that we learned about in chapter 8.

You probably know that carbohydrates, (also called oligosaccharides), are sugar polymers. Cellulose, the tough, fibrous material in plants, is one of the most abundant organic compounds found in nature. It is basically a long chain of glucose molecules linked together:



glucose

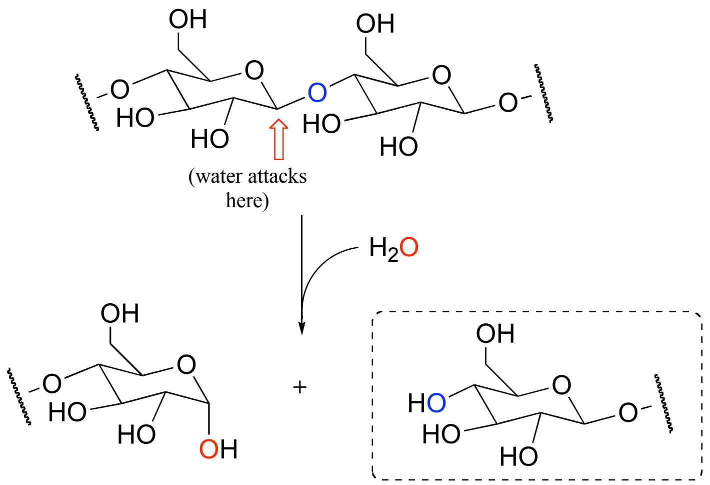


a glucose-glucose linkage in cellulose

The bonds connecting individual sugar units in carbohydrate chains are called **glycosidic bonds**, and enzymes that break them are called **glycosidases**. Humans do not have glycosidases capable of breaking the glycosidic bonds in cellulose - and thus we cannot use cellulose a source of energy - but cellulose is nonetheless necessary to us as dietary fiber. Cows and other ruminants are able to derive energy from cellulose because they maintain bacteria in their digestive tract which possess the proper glycosidase enzymes.

Starch, the carbohydrate that we eat in bread and pasta, is also a long chain of glucose molecules. However, the glycosidic bonds linking the individual glucose units in starch have a different stereochemical configuration from those in cellulose, and humans *do* have the proper glycosidase enzymes to digest starch. We will discuss carbohydrates again in chapter 11, and you will learn many more details if you take a course in biochemistry.

We focus now on the chemistry by which glycosidic bonds are broken and formed. Enzymes called glycosidases catalyze these reactions. Consider for example the following reaction, catalyzed by a bacterial enzyme, in which the glycosidic bond between two glucose molecules in cellulose is cleaved.



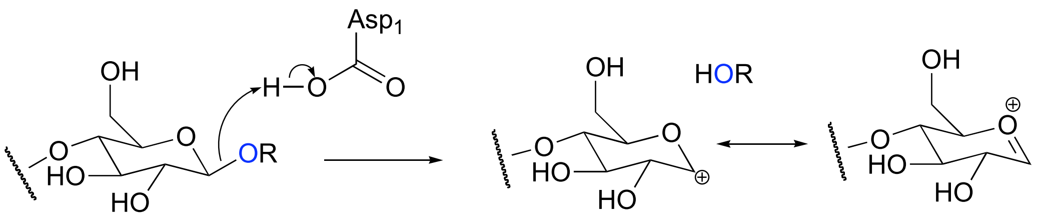
(water attacks here)

(= HOR)

This is a hydrolysis reaction – recall that this term is used to describe any reaction where a bond is being broken by water. If you look carefully, you will recognize that this reaction is simply a nucleophilic substitution at the carbon indicated by the arrow: water is the nucleophile, and the leaving group is an alcohol, specifically the glucose molecule on the right side. In the next few figures, the leaving group will be referred to as 'HO-R' for simplicity.

(A quick note on hydrolysis reactions: we will see many more examples of hydrolysis reactions throughout our study of organic chemistry. In eukaryotic cells, most hydrolysis reactions occur in the acidic (pH ~4.5) environment of the lysosome, an organelle that specializes in breaking large molecules down into small ones.)

Evidence suggests that glycosidase reactions probably occur through an S_N1 mechanism, implying the formation of a short-lived cationic intermediate. Here is the first, rate-determining step:

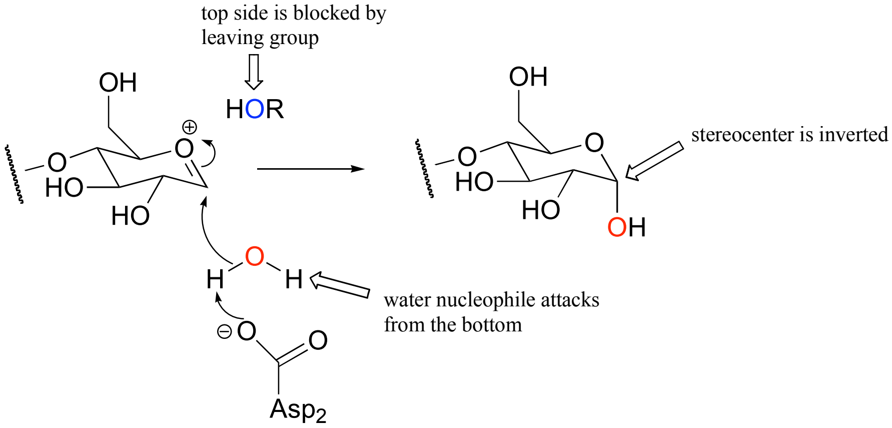


oxonium intermediate

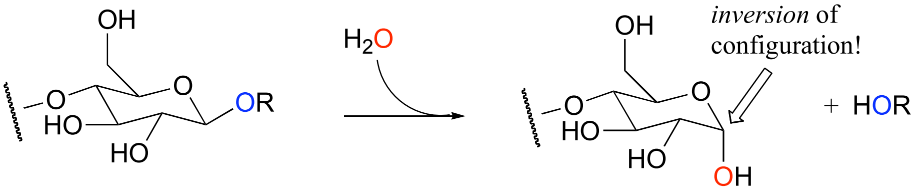
Notice that the positively charged carbon on the intermediate is adjacent to an oxygen. Recall from previous discussions (section 8.4B) that oxygen can act as a powerful electron *donating* group because of the resonance effect of its lone pairs. This is best illustrated by drawing a second resonance contributor, in which the positive charge is placed on the oxygen. This intermediate is generally referred to as an **oxonium ion**.

The active site of the enzyme has two aspartate residues, one positioned above the substrate (Asp$_1$) and one below (Asp$_2$). The leaving group is protonated by Asp$_1$: protonation, as you recall, creates a better leaving group.

Here is the second step of the glycosidase mechanism:


top side is blocked by leaving group
HOR
OH
O
HO
HO
H
O
H
O
O
Asp2
water nucleophile attacks from the bottom
stereocenter is inverted
OH

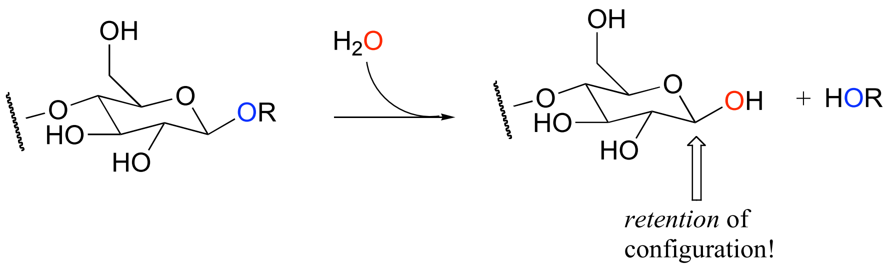
Notice that this substitution occurs with *inversion of configuration*:



This is because the leaving group remains bound in the active site after the formation of the oxonium ion intermediate, and blocks the bottom side of the electrophilic carbon from attack. The water nucleophile, as it attacks from the bottom side, is deprotonated by Asp$_2$.

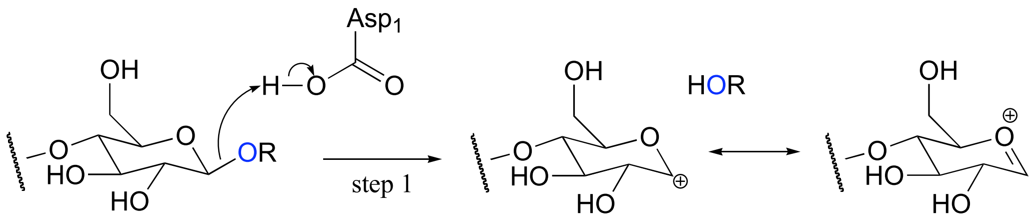
Because the reaction results in inversion of configuration, the enzyme is called an **inverting glycosidase**.

Researchers have also identified **retaining glycosidases**, which catalyze similar hydrolysis reactions except with *retention* of configuration:

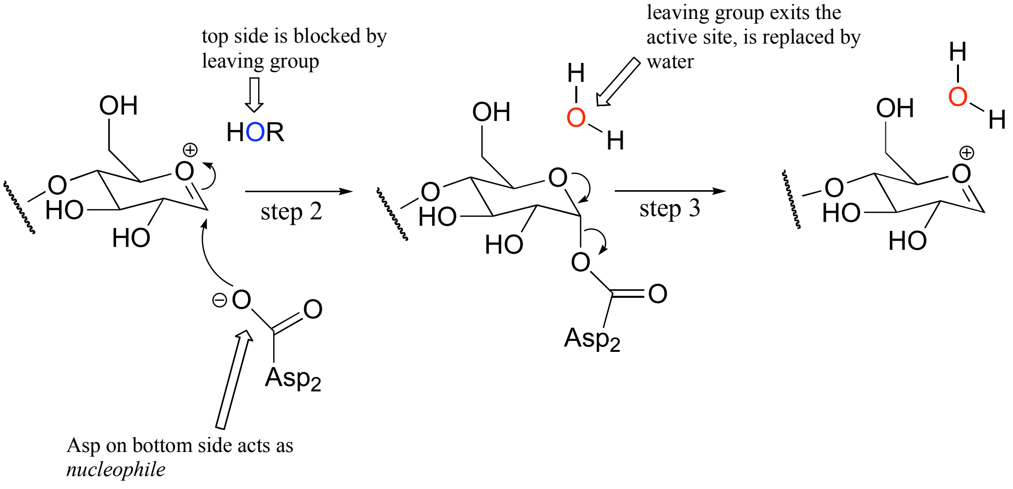


retention of configuration!

The active site architectures of the inverting and retaining glycosidases are actually very similar: both have two aspartate residues positioned above and below the electrophilic carbon. The first step in the retaining mechanism is the same as in the inverting mechanism: formation of the oxonium intermediate.



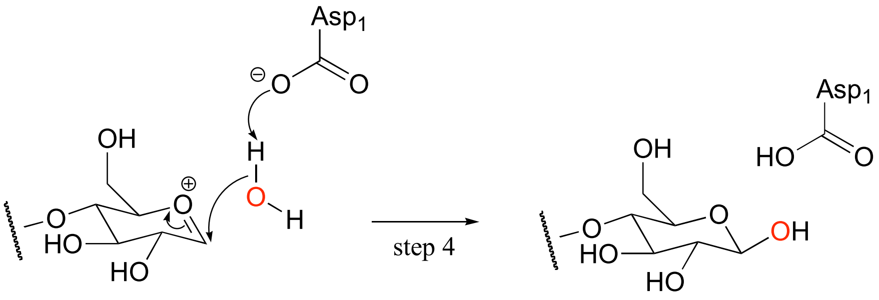
In the retaining reaction, however, the bottom side aspartate (Asp_2) acts as a *nucleophile* instead of as a base (step 2 below) , attacking the electrophilic carbon directly and forming what is referred to as a **covalent intermediate** (the enzyme is covalently attached to the terminal glucose molecule *via* the aspartate residue).


top side is blocked by leaving group

leaving group exits the active site, is replaced by water

Asp on bottom side acts as *nucleophile*

In the meantime, ROH (the other section of the celluose chain) diffuses away from the active site. The substrate-enzyme linkage then breaks (step 3), resulting in formation of the same oxonium ion intermediate as before. A nucleophilic water molecule then attacks from the *top* side, as Asp_1 serves as a catalytic base (step 4).



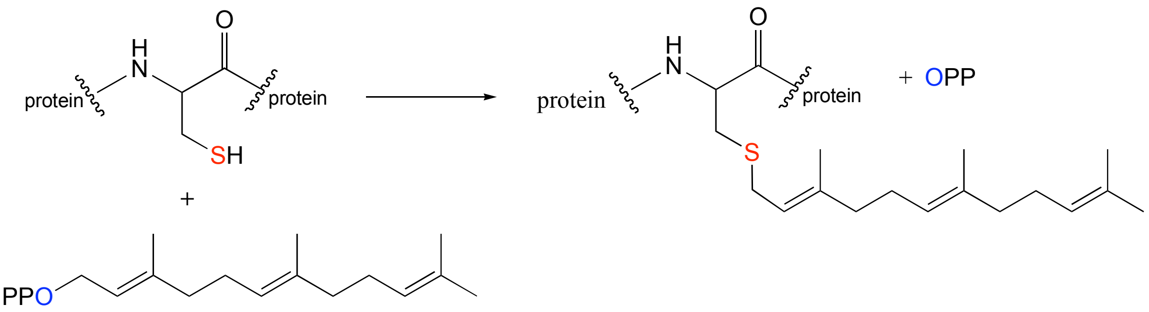
What has occurred here is known as a **double displacement** mechanism: in order to achieve substitution with retention of configuration, two nucleophilic substitutions have occurred in a row, each with inversion of stereochemistry. The overall result of two successive inversions at the same stereocenter is, of course, retention.

It is important to remember that the stereospecificity observed in the two examples of S_N1 reactions above is due to the exquisite control that enzymes have over the reactions they catalyze. Nucleophiles and electrophiles are bound in precise locations so that nucleophilic attack can occur from one side and one side only. While concerted S_N2 displacements take place specifically with inversion of configuration in both enzymatic and non-enzymatic situations, remember (section 8.2B) that nonenzymatic S_N1 reactions generally result in racemization. Recall also the reasoning behind this: with no enzymatic control, the nucleophile can attack the planar, sp^2-hybridized carbocation intermediate from either side with equal probability, resulting an a racemic mixture of both the R and S product.

Section 9.3: Protein prenyltransferase: a hybrid S_N1/S_N2 substitution

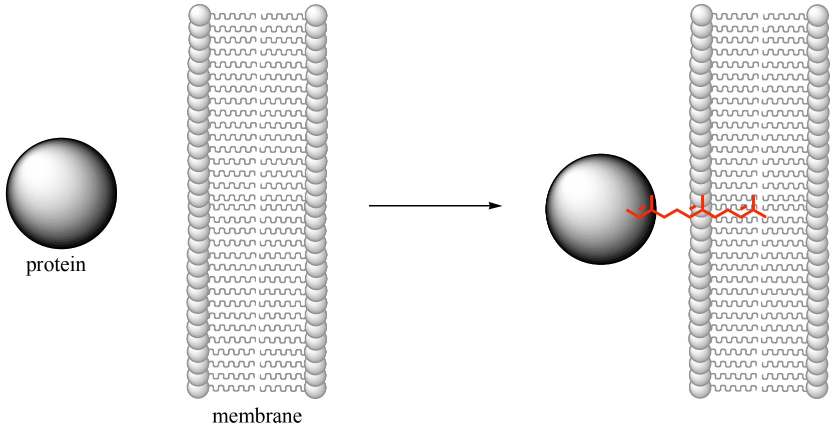
9.3A: The biological relevance of the protein prenyltransferase reaction

Although it is instructive to consider examples of nucleophilic substitution reactions that are thought to occur by completely S_N1 or completely S_N2 mechanisms, the reality is that many biologically relevant substitution reactions fall somewhere in between these two extremes. A good example of this is the reaction catalyzed by an enzyme called protein prenyltransferase:



In this nucleophilic substitution reaction, a cysteine residue on certain proteins is modified with a branched-chain 'isoprenyl' hydrocarbon. Do not be confused by the fact there are two proteins involved in this reaction - one is the catalyst (the enzyme, protein prenyltransferase), and the second is a *substrate* in the reaction.

This reaction, and the chemical mechanism involved, is of particular interest to biomedical scientists. Many of the proteins that are substrates for this type of modification are involved in cell signaling processes, and they are not able to carry out their biological functions unless they are *anchored* to a cell's lipid bilayer membrane. This is the role played by the isoprenyl group (see section 2.4B for a description of the molecular structure of lipid bilayer membranes).



unmodified protein cannot associate with membrane - biologically inactive

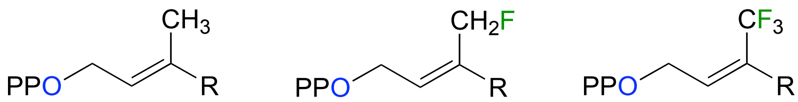
modified protein can anchor to membrane - biologically active

Some proteins that require this isoprenyl anchor attachment have been implicated in tumor formation. Scientists hope that if they can find a way to shut down the prenylation reaction, the tumor-causing proteins will not be able to anchor to cell membranes and thus will remain inactive. The search is on for an effective inhibitor of this enzyme to serve as a potential anti-tumor drug.

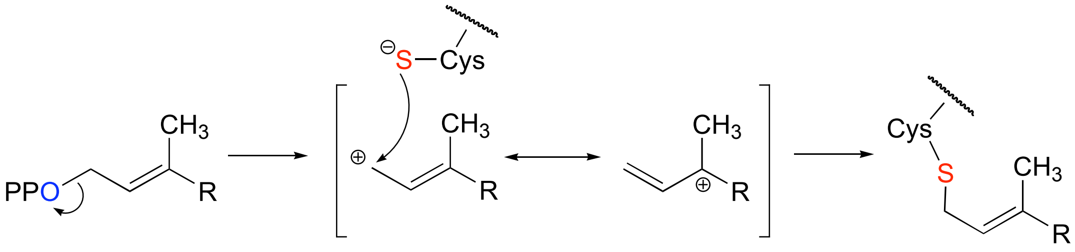
Experimental evidence (see section 9.3C) indicates that when the substrate protein is bound to the active site of the enzyme, the cysteine is in a deprotonated state (a thiolate ion). A thiolate is a *very* potent nucleophile! Notice that the leaving group in this reaction is inorganic pyrophosphate.

9.3B: Determining the mechanism of protein prenyltransferase with fluorinated substrate analogs

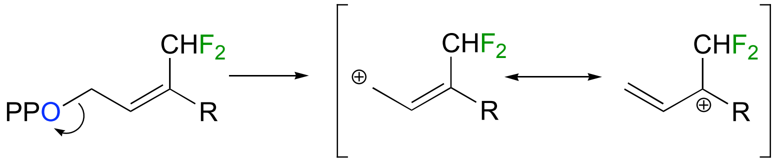
Scientists are very interested in understanding the details of the protein prenyltransferase reaction. One question that has been successfully addressed is whether the mechanism is S_N1, S_N2, or somewhere in between. Researchers measured the rate of the reaction with the normal isoprenyl substrate, and also with two synthetic derivatives in which nearby methyl hydrogens were replaced with either one or three fluorines (*Proc. Natl. Acad. Sci. U.S.A*. **1995**, *92*, 5008).



Fluorine has an atomic radius that is close to that of hydrogen, and thus a hydrogen-to-fluorine substitution does not effect the sterics of the reaction to a great extent. However, as you well know, fluorine is very different from hydrogen in terms of electronics: fluorine is highly electronegative (in other words, highly electron-withdrawing). If the reaction is S_N1, it should proceed through a positively-charged intermediate: this intermediate would be an allylic carbocation (see section 8.4B) , with the positive charge delocalized by resonance.



Fluorines, being powerful electron-withdrawing groups, would significantly destabilize this hypothetical carbocation intermediate, and thus would slow down the first (rate-determining) step of an S_N1 reaction.

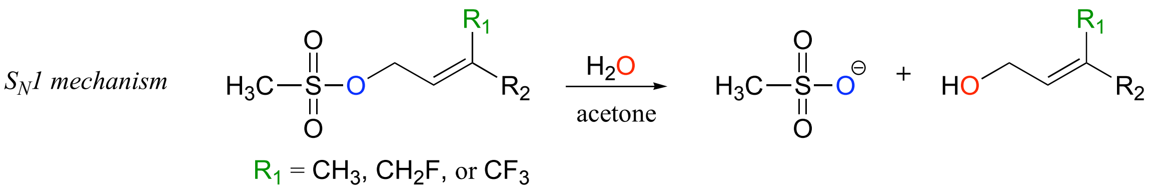


electron-withdrawing fluorines destabilize the intermediate, slow down the rate-determining step

Thus, it stands to reason that the overall rate of an S_N1 reaction would decrease each time a fluorine substitution is made, and that the slowest reaction rate would be with the trifluorinated substrate.

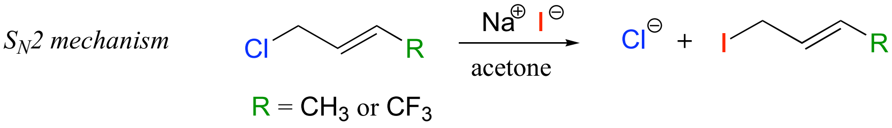
Conversely, if the reaction is S_N2, the fluorines should *not* slow down the rate of the reaction because there is no carbocation intermediate to destabilize.

This experimental premise is supported by similar experiments with model non-enzymatic substitution reactions that are known to proceed by S_N1 or S_N2 mechanisms. As expected for the S_N1 model reaction, adding fluorines slowed down the reaction considerably.

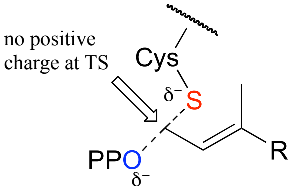


Relative to the compound where $R_1 = CH_3$, rates of reaction were 100,000 times slower for the monofluorinated substrate, and 10 million times slower for the trifluorinated substrate (*J. Am. Chem. Soc.* **1981**, *103*, 3926).

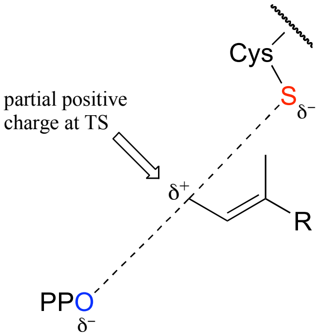
In the model S_N2 reaction, however, fluorine substitution on a slightly different substrate resulted in moderately *faster* rates of substitution (*Biochem*. **1977**, 16, 5470; *J. Am. Chem. Soc.* **1962**, *84*, 3157).



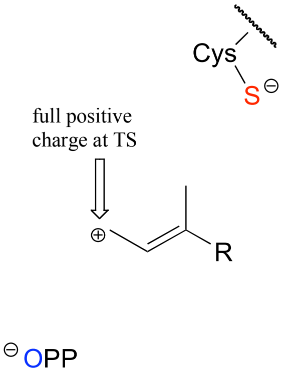
With the protein farnesyltransferase reaction, monofluorination of the substrate resulted in a 10-fold lower reaction rate, while the trifluorinated substrate analog reacted at a rate approximately 1000 times slower than the normal substrate. Clearly, the fluorine substitutions slowed the reaction down, but the observed rate reductions were not nearly as dramatic as those for the model S_N1 reaction. The researchers concluded that this reaction mechanism is a hybrid: mainly S_N2, but with *some* S_N1 character. In other words, a fully-charged carbocation intermediate does not form, but there is a **late transition state**: the bond between the electrophilic carbon and the oxygen of the pyrophosphate leaving group has come close to breaking by the time the thiolate approaches.



'pure' S_N2 transition state



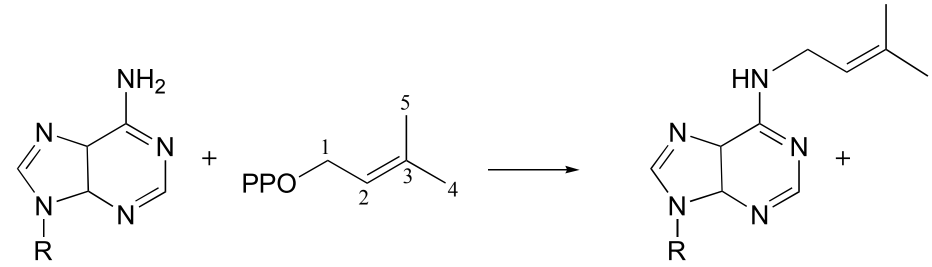
'late' transition state
(hybrid mechanism)



'pure' S_N1
intermediate

As the reaction proceeds from ground state to transition state, the partial positive charge on the electrophilic carbon increases, which explains the moderately decelerating effect of the electron-withdrawing fluorines.

Exercise 9.3: The enzyme AMP - DMAPP synthase catalyzes the following reaction:



Adenosine monophosphate (AMP)

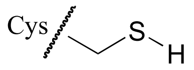
dimethylallyl diphosphate (DMAPP)

Would you expect fluorine substitution at carbon #5 of the DMAPP substrate to slow the rate of this reaction to a greater or lesser extent than what was observed for the protein farnesyltransferase reaction? Explain your reasoning.

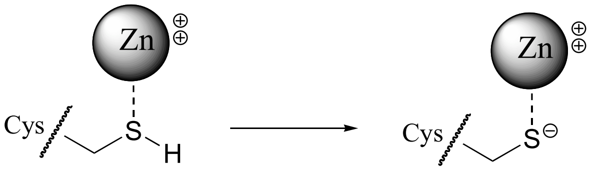
Exercise 9.4: Look again at the two 'model' S_N1 and S_N2 reactions referred to in the study described in this section. What aspects of each reaction would lead you to predict their mechanisms, if you had not been told?

9.3C: The zinc-thiolate interaction in protein prenyltransferase - 'tuning' the nucleophile

The nucleophile in the substitution reaction catalyzed by protein farnesyltransferase is thought to be a deprotonated thiolate ion (RS^-) rather than a thiol. Evidence for a thiolate nucleophile is based on UV spectroscopy experiments (*Biochemistry* **1999**, *38*, 13138). Thiolates, unlike thiols, have a characteristic UV absorbance at 236 nm. Absorbance at 236 nm was observed to increase upon binding of the protein substrate to the enzyme, indicating that as the protein substrate enters the enzyme's active site, the critical cysteine is deprotonated. When the substrate protein is free in solution, the pK_a of the cysteine thiol is determined experimentally to be approximately 8.5, meaning that it should be mainly protonated at neutral pH. The enzyme, through its interactions with the substrate protein, is able to lower the effective pK_a of the cysteine to the point where the thiolate ion is the predominant species. This is accomplished primarily by coordination of the cysteine sulfur atom to a Zn^{2+} ion that is bound in the active site of the enzyme: the positive charge on the zinc ion balances the negative charge of the cysteine thiolate.



free cysteine:
$pK_a \sim 8.5$

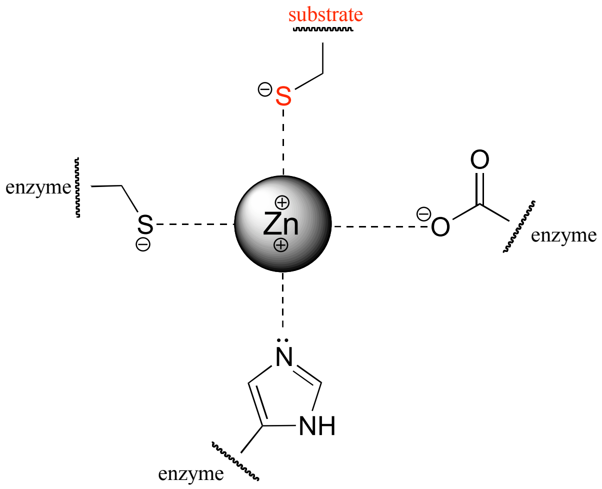


zinc-bound cysteine:
$pK_a < 5.5$

Evidence suggests that coordination to the zinc ion lowers the pK_a of the cysteine thiol by approximately 3 pK_a units.

This raises an intriguing question that is fundamental to our understanding of how enzymes catalyze reactions. The job of the zinc ion is to lower the pK_a of the substrate cysteine, and it does this by stabilizing the negative charge on the conjugate base (thiolate). However, if the zinc stabilizes the thiolate *too* much, the thiolate will be a less reactive - and thus less effective - nucleophile! If the zinc-thiolate interaction is too strong, in other words, the thiolate will not be able to dissociate from the metal, and nucleophilic attack will not occur. A delicate energetic balance between stability and reactivity is at work here: the zinc must provide just the right amount of charge stabilization to lower the pK_a and ensure deprotonation of the thiol, but not so much that the thiolate is no longer reactive enough as a nucleophile.

In effect, the overall positive charge of the zinc ion, and by extension its ability to stabilize the thiolate, must be 'tuned' to achieve the optimal degree of stabilization. How is this accomplished? The answer lies in the way that the zinc ion is specifically bound in the active site of the enzyme. In addition to its interaction with the *substrate* cysteine, the zinc ion is also coordinated to three amino acid residues from the *enzyme*: a cysteine, a histidine, and an aspartate.



These three ligands, to differing degrees, donate electron density to the zinc, making it less positive. By fine-tuning the overall degree of positive charge on the zinc ion, the enzyme is able to control the extent to which the negative charge on the substrate thiolate is balanced, ensuring that the cysteine pK_a is lowered just enough to ensure deprotonation without compromising nucleophilicity.

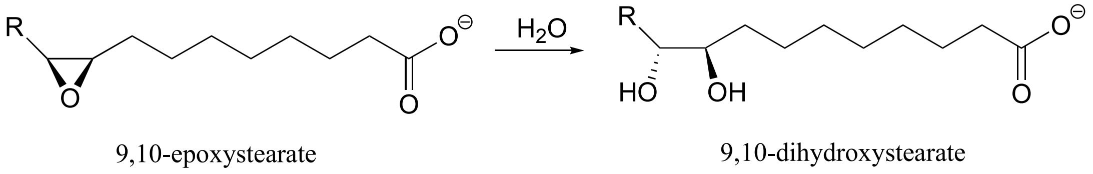
Researchers studying this phenomenon used site-directed mutagenesis (section 6.5C) to change both of the negatively-charged zinc-coordinating residues (aspartate and cysteine) to neutral histidines, resulting in a zinc center with a substantially greater overall positive charge (*Biochem.* **2002**, *41*, 10554). While this mutant enzyme was able to bind zinc and the two substrates almost as well as the wild-type, *no enzymatic activity was detected*. It is likely that the substrate cysteine is deprotonated upon binding to the enzyme, but the resulting zinc-thiolate bond is so strong (due to the greater positive charge in the zinc ion) that the thiolate is unable to break away and attack the electrophilic carbon of the farnesyl diphosphate substrate. In effect, the thiolate is 'stuck' on the zinc, and thus is not able to act as a nucleophile.

If you look back at section 8.5C and think about the reasons why synthetic chemists use polar aprotic solvents in laboratory S_N2 reactions, you'll see that many of the same ideas are in play.

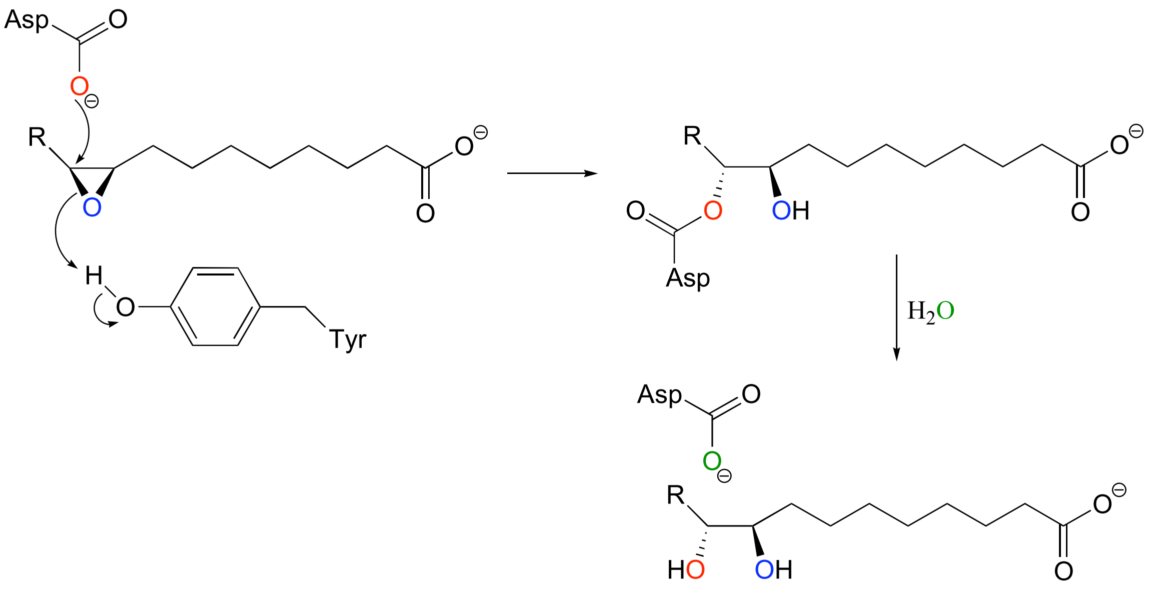
Section 9.4: Biochemical nucleophilic substitutions with epoxide electrophiles

9.4A: Hydrolysis of stearic acid epoxide: investigating the mechanism with kinetic isotope effect experiments

In chapter 8 we discussed some examples of nonenzymatic epoxide ring-opening reactions, and how they can proceed with either S_N1-like or S_N2-like mechanisms, depending on whether the reaction occurs in acidic or basic conditions. In a biological context, enzymes which catalyze the hydrolytic ring-opening of epoxides are called epoxide hydrolases. An epoxide hydrolase that has been isolated from soybeans, for example, is able to catalyze the hydrolysis of an epoxide derivative of stearate, a common fatty acid (*J. Biol. Chem.* **2005**, *280*, 6479).



For a long time, scientists thought that epoxide hydrolase reactions involved direct nucleophilic attack by water. More recently, however, it has been determined that the ring-opening nucleophile is actually an aspartate in the enzyme's active site.



An active site acid (a tyrosine in this particular enzyme) protonates the epoxide oxygen, which lowers the energy barrier for ring-opening.

(You may be a little confused by what is happening in the second step, when the water molecule enters the picture. Don't worry about this yet - this type of reaction is the subject of chapter 12, and will make perfect sense when we get there. If you are looking at a color version of the figure above, however, you will notice from the color scheme that the oxygen in water does *not* end up in the diol product.)

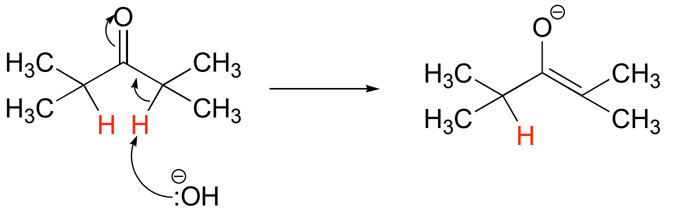
Exercise 9.5: The second step of this reaction could be envisioned as a nucleophilic substitution, with water as the nucleophile and aspartate as the leaving group. In fact, this would be very similar to the mechanism of the retaining glycosidases covered previously. However, even if the oxygens in the figure above were not colored, the stereochemical outcome of the reaction should tell you that the second step is *not* an S_N1 or S_N2 reaction. Explain. (You will revisit this chemical step in a problem in chapter 12)

Let's focus on the first step, which is a nucleophilic substitution / epoxide ring opening reaction. Is this first step S_N1-like or S_N2-like? Researchers studying other epoxide hydrolase enzymes have presented evidence for a concerted, S_N2-like mechanism. However, this particular plant enzyme appears to operate by a more S_N1-like mechanism. What is the evidence that suggests this?

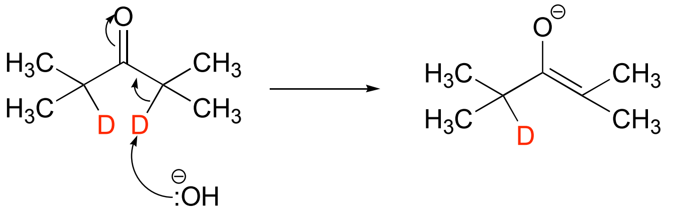
Previously, we discussed an experiment which used fluorine substituents on the reaction substrate to address this S_N1 vs S_N2 question. In this case, a very different approach was taken, using a phenomenon called the **kinetic isotope effect**.

When a hydrogen atom on a reacting molecule is replaced by a heavier deuterium or tritium isotope, small but measurable effects on the rate of the reaction are sometimes observed. These are called ‘kinetic isotope effects’ (KIE's). The term can also be applied to other isotopic substitutions such as ^{12}C to ^{14}C, but KIE's are most pronounced for hydrogen, because its heavy isotopes are two and three times heavier than ^{1}H.

When the α-hydrogens of a ketone are replaced by deuterium, for example, the rate constant (k) for the acid-base reaction decreases by a factor of six (*J. Am. Chem. Soc.* **1972**, *94*, 8351).



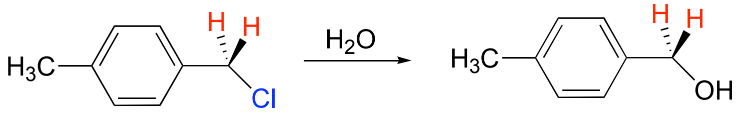
relative k_H = 6.1



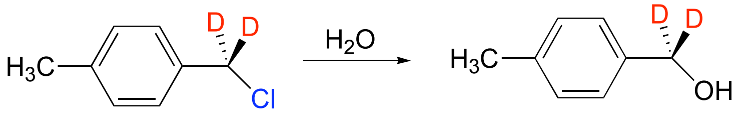
relative k_D = 1.0

The KIE is expressed as the ratio of rate constants k_H/k_D (or k_H/k_T if tritium is used). In the above example, k_H/k_D = 6.1, meaning that replacement with deuterium slows down the reaction by a factor of 6.1 times. The reasoning behind the KIE effect is somewhat complex, but suffice it to say that the heavier mass of the deuterium makes the activation energy for the bond-breaking step higher. In the acid-base example above, the hydrogen/deuterium is *directly involved* in the bond-breaking event: it is being abstracted by a base. The observed effect is thus called a **primary KIE**.

A KIE is also often observed when the hydrogen isotope is not *directly* involved in the reaction, but rather is bound to a carbon that is involved in a bond-breaking or bond-forming step. This is called a **secondary KIE**, and an example is shown below: (*J. Am. Chem. Soc.* **1970**, *92*, 232).



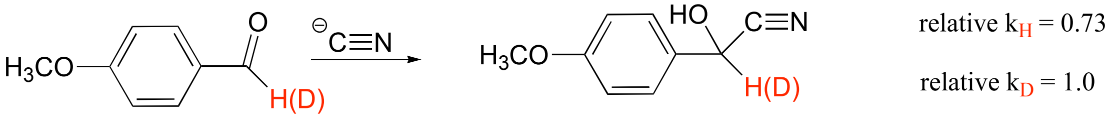
relative k_H = 1.3



relative k_D = 1.0

A secondary KIE can be either 'normal' ($k_H/k_D > 1$) or 'inverse' ($k_H/k_D < 1$, where the reaction with the heavier isotope actually goes *faster*).

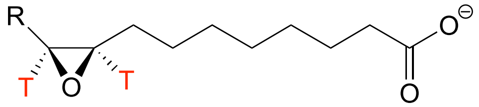
The secondary KIE can be useful in determining reaction mechanism. In general, normal secondary KIE's are observed when the rate-determining step involves a carbon center changing from sp^3 to sp^2 hybridization, in particular when this step involves the formation of a carbocation. An S_N1 reaction is expected to show a normal secondary KIE, because in the rate-determining step the carbon changes from sp^3 neutral charge to sp^2 positive charge. The reaction shown above, which is an S_N1 hydrolysis, has a 'normal' $k_H/k_D = 1.3$. Conversely, if the rate-determining step involves a carbon center changing from sp^2 to sp^3 hybridization, an inverse secondary KIE is often observed (we will study this reaction type in chapter 11).



(*J. Am. Chem. Soc.* **1972** *94*, 7579)

In S_N2 substitutions, where there is no sp^3 to sp^2 change in hybridization and no carbocation intermediate, the KIE is generally either slightly inverse or close to unity ($k_H/k_D = 1$).

For the soybean epoxide hydrolase reaction described earlier, a normal secondary KIE of $k_H/k_T = 1.30$ was observed when a tritiated substrate was used (*J. Biol. Chem.* **2005**, *280*, 6479).



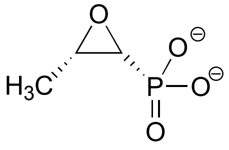
This indicates (but is not by any means conclusive proof of) a S_N1-like mechanism for the ring-opening step of the reaction.

Note: The non-enzymatic KIE examples in this section were taken from Advanced Organic Chemistry by Francis A. Carey and Richard J. Sundberg, which is recommended as an excellent source for more information on the kinetic isotope effect (and many other topics in organic chemistry. Many more examples of how KIE experiments can be used to investigate enzyme mechanisms can be found in the references below:

1) Cook, P. F. (1991) in *Enzyme Mechanism from Isotope Effects* (Cook, P. F., ed) CRC Press, Boca Raton, FL
2) Schramm, V. L. *Acc. Chem. Res.* **2003** *36*, 588.

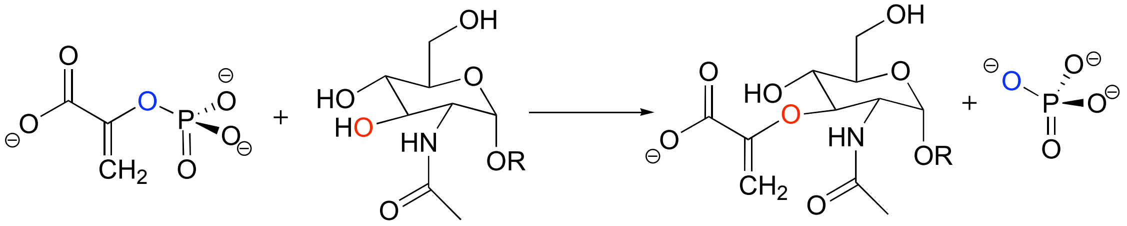
9.4B: Fosfomycin: an epoxide antibiotic

Fosfomycin is an antibiotic drug that acts by irreversibly inhibiting an enzyme involved in the biosynthesis of bacterial cell walls.



fosfomycin

Specifically, fosfomycin shuts down the enzyme UDP-*N*-acetylglucosamine enolpyruvyl transferase, commonly known as MurA, which catalyzes a critical step in the construction of bacterial cell walls.

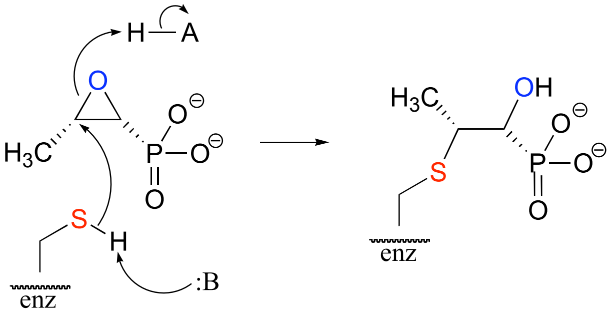


phosphoenolpyruvate (PEP)

UDP-*N*-acetylglucosamine

This may appear to be simply an S_N1 or S_N2 reaction - but recall (section 8.2E) that nucleophilic substitutions generally do not occur at alkene carbons! The enzymatic reaction itself is somewhat complex and outside the scope of the present discussion, but the important point for now is that an active site cysteine is known to be essential for the enzyme's catalytic function.

The key to the action of fosfomycin is its reactive epoxide group (*Biochem*. **1994**, *33*, 10,646), which serves as an electrophilic 'bait' for the active site cysteine.

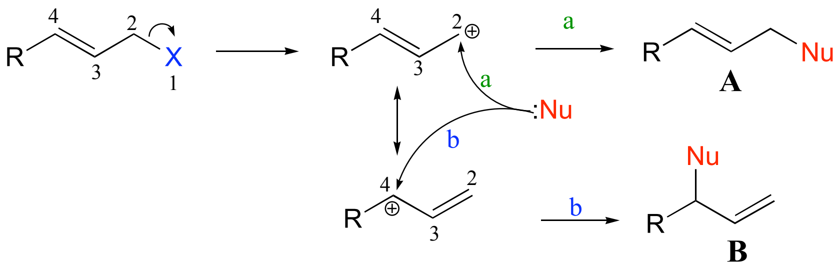


When the fosfomycin molecule (which is structurally a mimic of the phosphoenolpyruvate substrate) enters the active site of the enzyme, the cysteine thiol attacks one of the electrophilic carbons and opens the ring, leading to the irreversible alkylation of the active site. This effectively shuts down the enzyme, leading to death of the cell.

> Exercise 9.6: Account for the regiochemical outcome of the enzyme-alkylating reaction shown above.

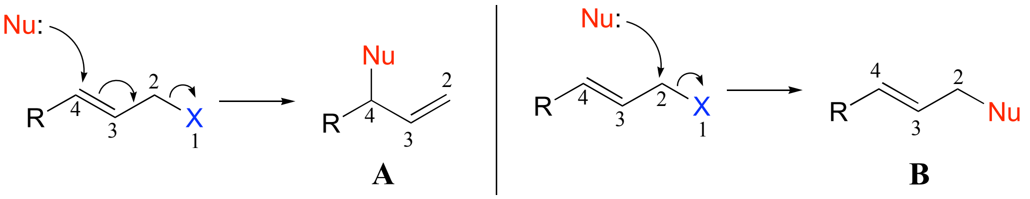
Section 9.5: Nucleophilic substitution over conjugated pi systems - the SN' mechanism

Until now, all of the nucleophilic substitution reactions that we have seen have involved *direct* displacement at a carbon center: that is, the carbon that is bonded directly to the leaving group is the target for the attacking nucleophile. However, this is not always the case: when the leaving group is bound to a carbon that is part of a system of conjugated π bonds, nucleophilic attack can occur at more than one site. Consider a generalized electrophilic species with a leaving group (X) in the allylic position. In an S_N1 mechanism, the carbocation intermediate would have two possible resonance forms.



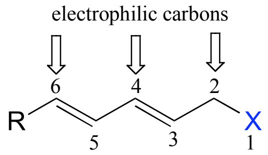
The incoming nucleophile, therefore, could attack either carbon 2 (the carbon originally bonded to the leaving group), or alternatively at carbon 4. These two events would lead to two different substitution products A and B, respectively.

The same can be said of an associative mechanism that does not involve a carbocation intermediate: in this case, however, the concerted mechanisms would be depicted as shown below:

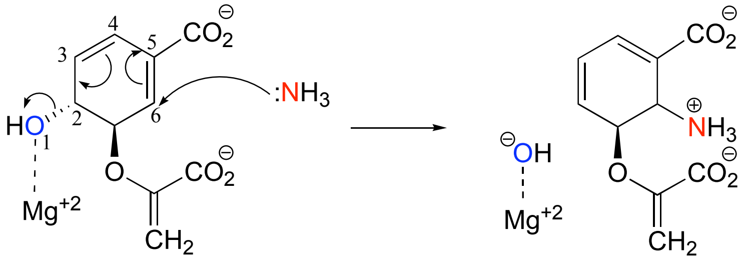


The mechanisms leading to product B are referred to as S_N', or nucleophilic allylic substitution. They are also sometimes referred to as 1,4 substitutions, according to a numbering system where the leaving group is designated atom #1.

If the conjugated π orbital system extends further, more S_N' products can result: in the following example, 1,6 as well as 1,4 and 1,2 substitutions are possible.

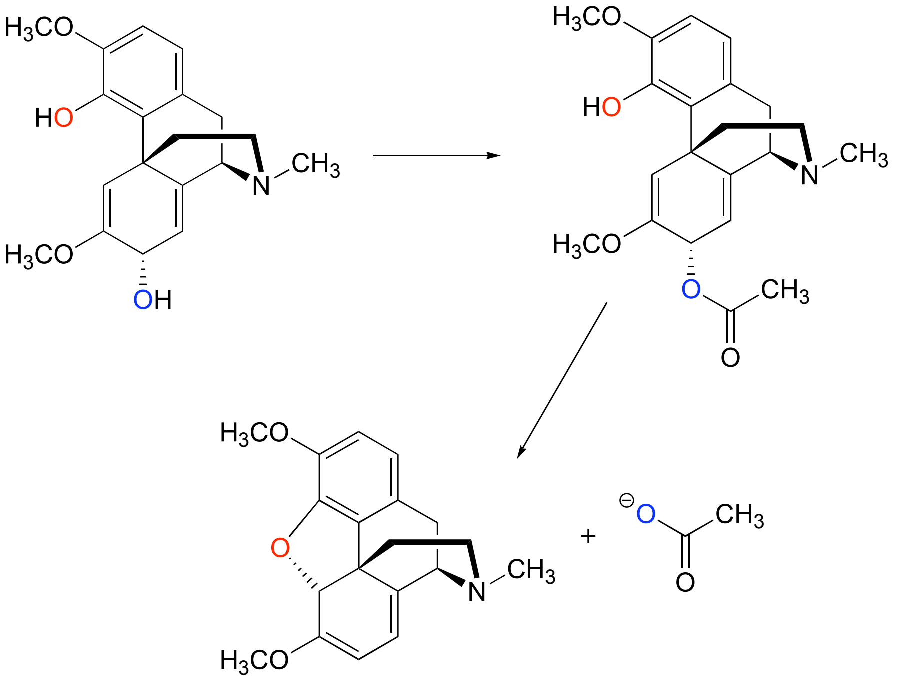


An important example of an enzymatic S_N2' reaction is the first step of the reaction catalyzed by anthranilate synthase, an enzyme involved in tryptophan biosynthesis.

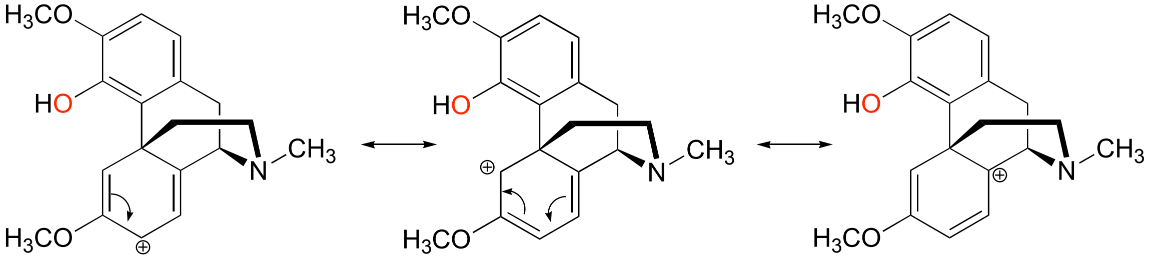


Nucleophilic ammonia attacks in a 1,6 fashion, displacing a hydroxide on the other side of the ring. Note that a hydroxide ion is normally a very poor leaving group. However, in this case the negative charge is stabilized by coordination to a magnesium ion (a Lewis acid) bound in the active site.

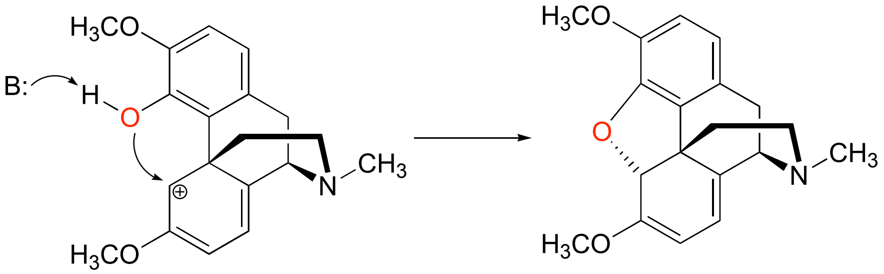
In the biosynthesis of morphine in opium poppies (*Science* **1967**, *155*, 170; *J. Biol. Chem* **1995**, *270*, 31091), the following 1,4 SN' ring-closing reaction occurs:



Notice that before the SN' cyclization takes place, the alcohol is first converted to an ester, which is a much better leaving group (we'll study this type of reaction in Chapter 12). Also, notice that the cyclization reaction itself might reasonably be expected to be stepwise (SN1') rather than concerted (SN2'), because the carbocation intermediate formed in a stepwise reaction would be stabilized very well by resonance delocalization of the positive charge:



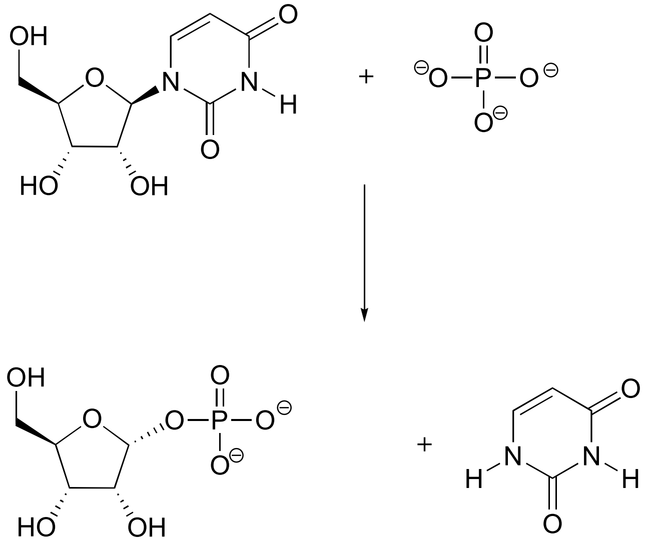
In addition, it is interesting to note that the cyclization stage of the reaction is spontaneous - in other words, it occurs without the need for enzyme catalysis, as soon as the alcohol is converted to an ester (making it a good leaving group).



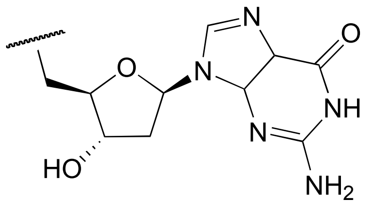
This is possible because the fused-ring structure of the molecule serves to hold the phenol nucleophile in the correct position for attack, thus greatly reducing the *entropic* component of the energy barrier.

Chapter 9 problems

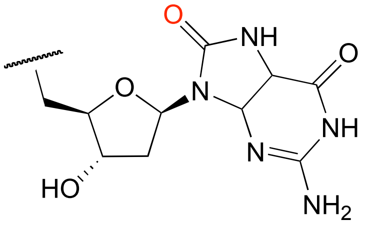
P9.1: Draw a complete curved arrow mechanism for the following reaction, which is the first step in the degradation of uridine. Assume an S_N1 mechanism, and that the leaving group leaves as an anion, then is protonated in a later step. Show how both the leaving group and the carbocation intermediate are stabilized by resonance.



P9.2: The guanine base of DNA can be oxidized to 8-oxoguanine by the so-called 'reactive oxygen species' (ROS) that are constantly being produced in our bodies as a byproduct of respiration (ROS are discussed briefly in section 17.2, as well as in most biochemistry textbooks).



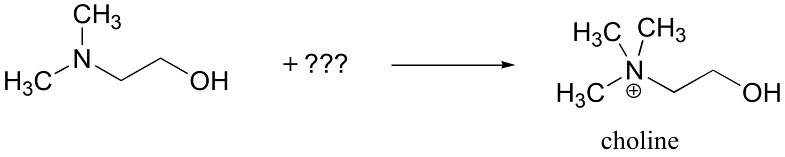
DNA guanine base



damaged 8-oxoguanine base

One way that our cells deal with this damaged DNA is to first remove the 8-oxoG base from the sugar-phosphate DNA chain. This is thought to occur with the help of a nucleophilic lysine residue in a type of DNA repair enzyme called DNA glycosylase, a reaction that results in a temporary enzyme-DNA covalent bond. Propose a likely a mechanism for this reaction.

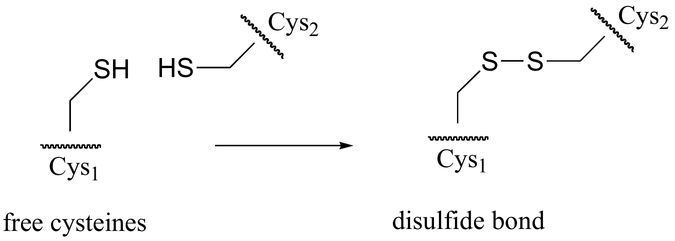
P9.3: Choline, an important neutotransmitter in the nervous system, is formed directly from 2-(N,N-Dimethylamino)ethanol:



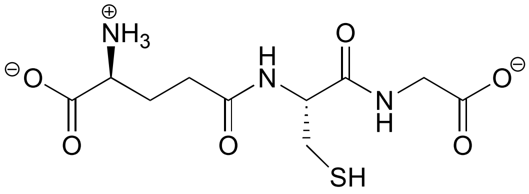
a) Besides the enzyme and the starting compound, what other important biomolecule do you expect to play a part in the reaction?

b) Show a mechanism for the reaction.

P9.4: In many extracellular proteins (proteins that are located outside cells) , the sulfur atoms of two cysteine side chains become covalently linked to form what are called 'disulfide bonds' (these bonds will be discussed in more detail in section 16.12)



Disulfide linkages are converted back to free cysteines in living cells by the action of a molecule called glutathione.

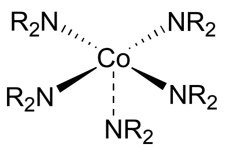


glutathione

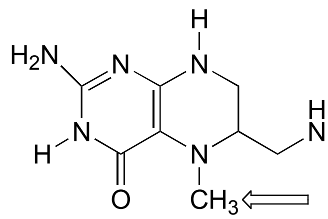
In a spontaneous (uncatalyzed) reaction, the cysteine-cysteine disulfide bond in the protein is broken, and a new disulfide bond forms between two glutathione molecules. This process involves two successive S_N2-like reactions (although the electrophiles are *not* carbons).

a) What do you predict is the key functional group in the action of glutathione?
b) Using appropriate abbreviations, propose a mechanism for this process.

P9.5: The final step in the biosynthesis of the amino acid methionine is dependent on two coenzymes: cobalamin, which is derived from vitamin B_{12}, and 5-methyltetrahydrofolate, a derivative of folic acid. Cobalamine is a large and complex molecule, containing a cobalt metal center that is bound to 5 nitrogen-containing groups.



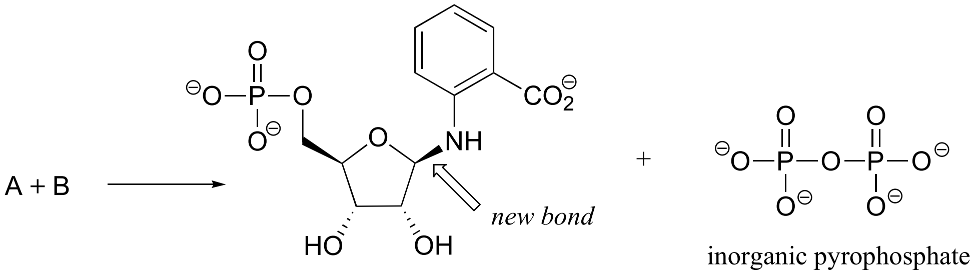
cobalamine
(just the metal center)



5-methyltetrahydrofolate

The cobalt metal in cobalamin is an excellent nucleophile. In the methionine biosynthesis reaction, the 5-methyl group from 5-methyltetrahydrofolate ends up on methionine, with overall *retention* of stereochemical configuration about the methyl carbon (this can be demonstrated by using ^{2}H- and ^{3}H-labels on the methyl group- see section 9.1A).. Propose a likely mechanism for this process, taking into account the stereochemical outcome (you will need to infer the precursor to methionine).

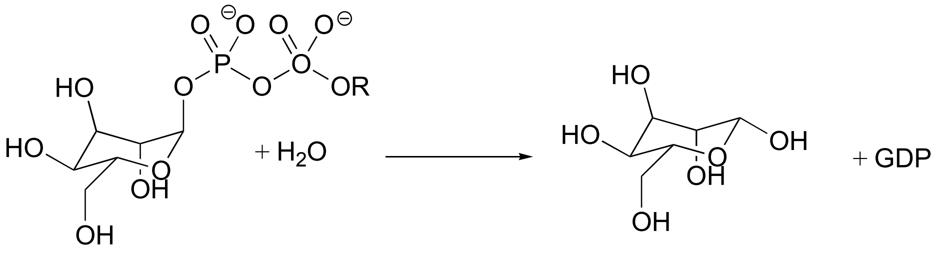
P9.6: The biosynthesis of tryptophan involves the formation of the compound whose structure is shown below. The new bond formed in the reaction is indicated by the arrow, and the reaction proceeds with retention of configuration at the anomeric carbon. Inorganic pyrophosphate is released. Show the structures of the two starting compounds A and B.



inorganic pyrophosphate

Challenge problems

C9.1: An enzyme found in bacteria catalyzes the hydrolysis of mannose sugar-phosphate (the phosphate group is actually GDP).



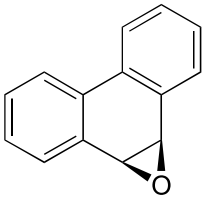
X-ray crystal structure of the enzyme's active site shows an aspartate residue located close to the anomeric carbon of the bound substrate. Site-directed mutagenesis of this aspartate to asparagine leads to a ($10^{2.6}$)-fold decrease in the maximal rate of the reaction. In addition, the crystal structure shows that the diphosphate of GDP interacts with three active site arginines, a tyrosine, and a bound magnesium ion (Mg^{+2}) is required for activity.

a) Does this information, taken together, lead you to predict the mechanism to be more dissociative or more associative in character? Explain your prediction.

b) Suggest a fluorine substituted substrate analog that could be used to further address the associative/dissociative character of the reaction.

c) How might using D_2O in the reaction buffer further clarify the situation?

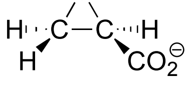
C9.2: Glutathione S-transferases (GSTs) are a family of enzymes that use the multifunctional coenzyme glutathione to break down potentially harmful foreign (often referred to as 'xenobiotic') compounds in the cell through nucleophilic attack. Predict the likely product(s) of the GST-catalyzed reaction between glutathione and the aromatic epoxide shown below, and show the relevant mechanism.



C9.3: A glutathione S-transferase enzyme is thought to be involved in conferring resistance in some bacteria to the antibiotic fosfomycin.

a) Draw a mechanism showing how this GST enzyme could inactivate fosfomycin.
b) The enzyme requires Mg^{+2} for activity. Propose two possible catalytic roles for the metal ion.

C9.4: Many bacteria have 'dehalogenase' enzymes, which are able to remove halogens from halogenated organic pollutants. This property makes them potentially very useful in environmental cleanup efforts. In studying the mechanism of one such enzyme, scientists synthesized an irreversible inhibitor with the specific structure shown below (the enantiomer was *not* an inhibitor).

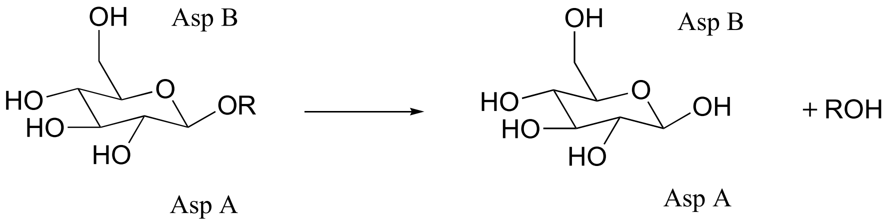


Mass spectroscopy revealed that the enzyme, after treatment with this compound, is alkylated at the carboxy-terminal proline residue.

a) Draw two possible enzyme inactivation mechanisms showing two possible proline alkylation products. Explain why your two mechanisms lead to two different products.

b) What do the findings of this study suggest about the mechanism of dehalogenation? Just state your answer in words.

C9.5: Glucosidase enzymes catalyze the cleavage of glucose-glucose bonds (R in the figure below is a second glucose molecule). Many glucosidases have two critical Asp residues in the active site, positioned above and below the site of hydrolysis:

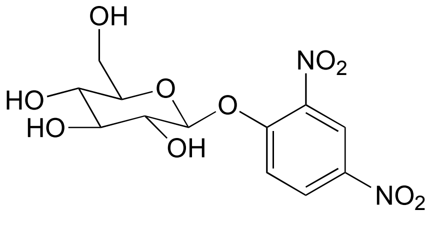


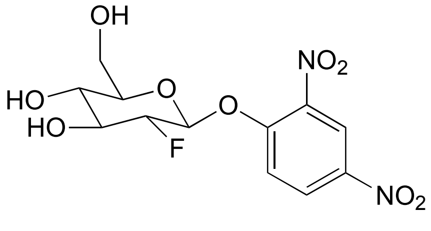
When Asp A was mutated to alanine, no enzymatic activity was observed when the enzyme was incubated with the substrate. When azide ion was added to the buffer, however, an azido-glucose product was observed (azide, N_3^-, is an excellent nucleophile).

a) Propose a mechanism for the normal hydrolysis reaction (with wild-type enzyme), including the role played by the two asparates.

b) Explain the formation of, and predict the structure of the azido-glucose product.

c) The substrate analogs shown below were tested with glucosidase. Why do you think they chose the dinitrophenol group in these molecules? What do you think researchers might have been trying to learn with this experiment?





Tables

List of tables:

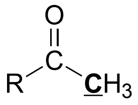
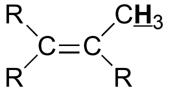
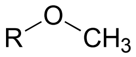
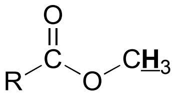
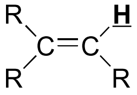
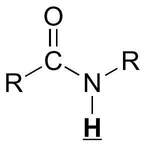
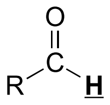
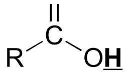
Table 1: Some characteristic absorption frequencies in IR spectroscopy
Table 2: Typical values for ^{1}H-NMR chemical shifts
Table 3: Typical values for ^{13}C-NMR chemical shifts
Table 4: Typical coupling constants in NMR
Table 5: The 20 common amino acids
Table 6: Structures of common coenzymes
Table 7: Representative acid constants
Table 8: Some common laboratory solvents, acids, and bases
Table 9: Examples of common functional groups in organic chemistry

Table 1: Some characteristic absorption frequencies in IR spectroscopy

Bond type	frequency (cm^{-1})
C-H alkanes	2950 – 2850
C-H alkenes	3080 – 3020
C-H aldehyde	~2900
C-H alkyne	~3300
alkyne triple bond	2250 – 2100 (s)
alkene double bond	1680 - 1620(s)
carbonyl, ketone	1725 – 1700 (s)
carbonyl, aldehyde	1740 – 1720 (s)
carbonyl, ester	1750 – 1730 (s)
carbonyl, acid	1725 – 1700 (s)
carbonyl, amide	1690 – 1650 (s)
O-H, alcohols	3600 – 3200 (s, broad)
O-H, acids	3000 – 2500 (broad)
C-O, alcohols, esters, ethers	1300 - 1000

s = strong absorbance

Table 2: Typical values for ^{1}H-NMR chemical shifts

Hydrogen type	**Chemical shift (ppm)**
$RC\underline{\mathbf{H}}_3$	0.9 - 1.0
$RC\underline{\mathbf{H}}_2R$	1.2 - 1.7
$R_3C\underline{\mathbf{H}}$	1.5 – 2.0
	2.0 – 2.3
	1.5 – 1.8
$RN\underline{\mathbf{H}}_2$	1 - 3
$ArC\underline{\mathbf{H}}_3$	2.2 – 2.4
$R-C\equiv C-\underline{\mathbf{H}}$	2.3 – 3.0
	3.7 – 3.9
	3.7 – 3.9
$RO\underline{\mathbf{H}}$	1 - 5
	3.7 – 6.5
	5 - 9
$Ar\underline{\mathbf{H}}$	6.0 – 8.7
	9.5 – 10.0
	10 - 13

Chemical shift values are in parts per million (ppm) relative to tetramethylsilane.

Table 3: Typical values for ^{13}C-NMR chemical shifts

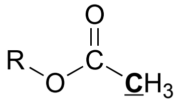
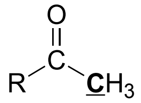
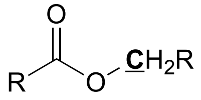
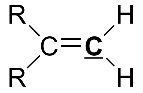
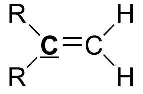
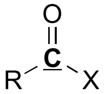
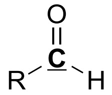
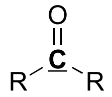
Carbon type	**Chemical shift (ppm)**
$R\underline{\mathbf{C}}H_3$	13 - 16
$R\underline{\mathbf{C}}H_2R$	16 - 25
$R_3\underline{\mathbf{C}}H$	25 - 35
	18 - 22
	28 - 32
$R\underline{\mathbf{C}}H_2NHR$	35 - 45
$R\underline{\mathbf{C}}H_2OH$	50 - 65
$R-C\equiv\underline{\mathbf{C}}-R$	65 - 70
$RO\underline{\mathbf{C}}H_2R$	50 - 75
	50 - 75
	115 - 120
	125 - 140
aromatic carbon	125 - 150
 (carboxylic acid derivatives)	165 - 185
	190 - 200
	200 - 220

Table 4: Typical coupling constants in NMR

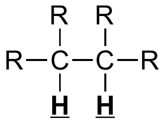
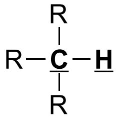
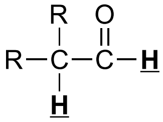
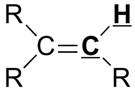
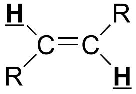
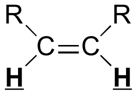
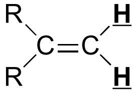
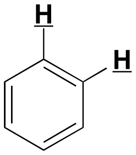
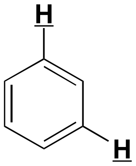
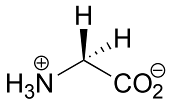
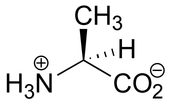
H-H coupling	J (Hz)	C-H coupling	J (Hz)
	6 -8		125 - 130
	2 - 3		150 - 170
	12 - 18		
	6 - 12		
	0 - 2		
	6 - 10		
	1- 3		

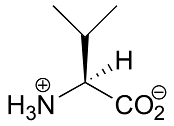
Table 5: The 20 common amino acids



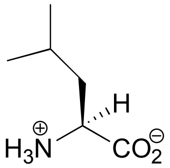
Glycine
(Gly, G)



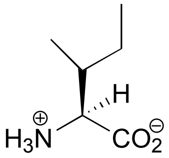
Alanine
(Ala, A)



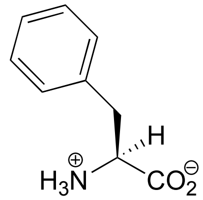
Valine
(Val, V)



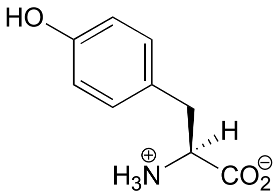
Leucine
(Leu, L)



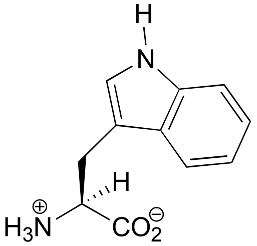
Isoleucine
(Ile, I)



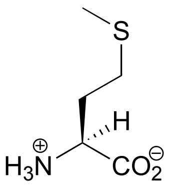
Phenylalanine
(Phe, F)



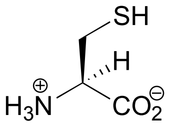
Tyrosine
(Tyr, Y)



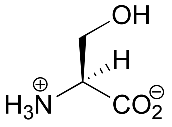
Tryptophan
(Trp, W)



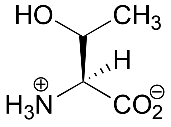
Methionine
(Met, M)



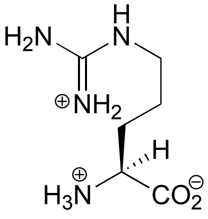
Cysteine
(Cys, C)



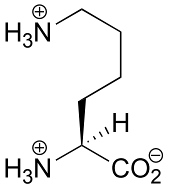
Serine
(Ser, S)



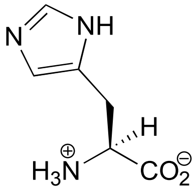
Threonine
(Thr, T)



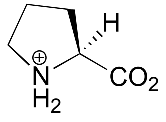
Arginine
(Arg, R)



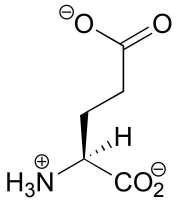
Lysine
(Lys, K)



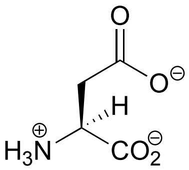
Histidine
(His, H)



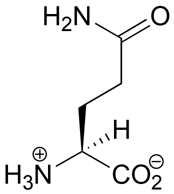
Proline
(Pro, P)



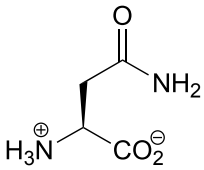
Glutamate
(Glu, E)



Aspartate
(Asp, D)

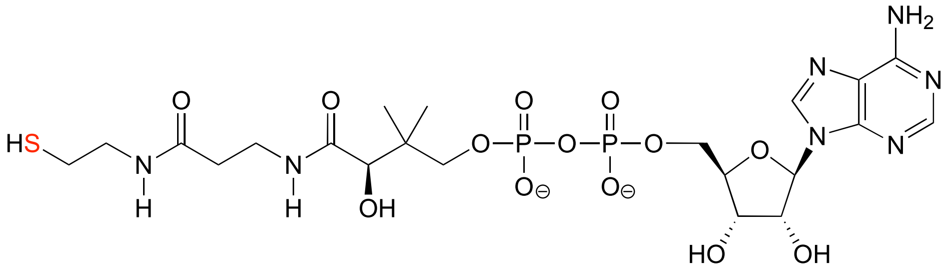


Glutamine
(Gln, Q)

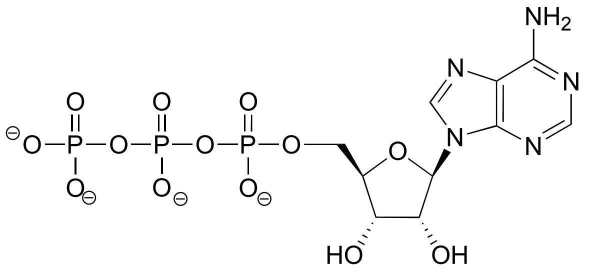


Asparagine
(Asn, N)

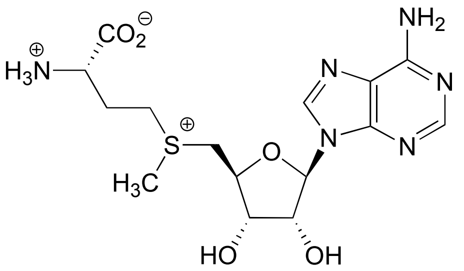
Table 6: Structures of common coenzymes



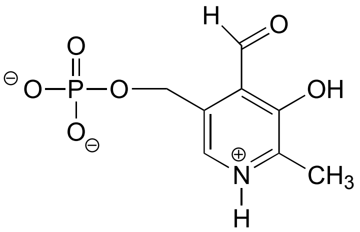
Coenzyme A
(HSCoA)



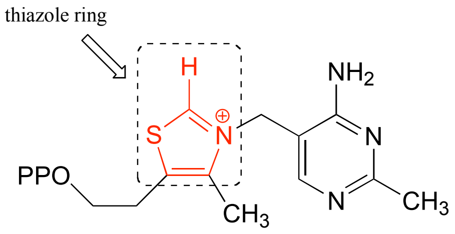
adenosine triphosphate
(ATP)



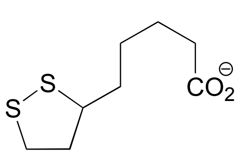
S-adenosyl methionine
(SAM)



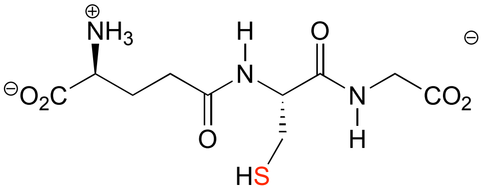
pyridoxal phosphate
(PLP)



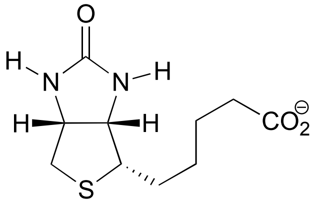
thiamine diphosphate
TPP)



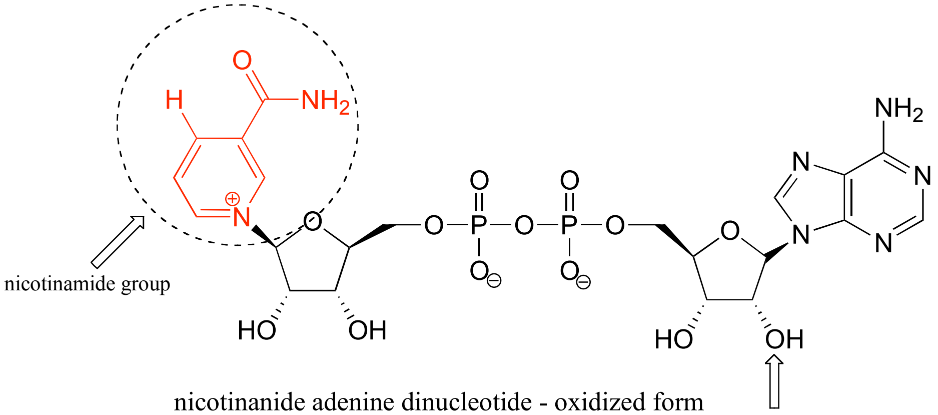
lipoate



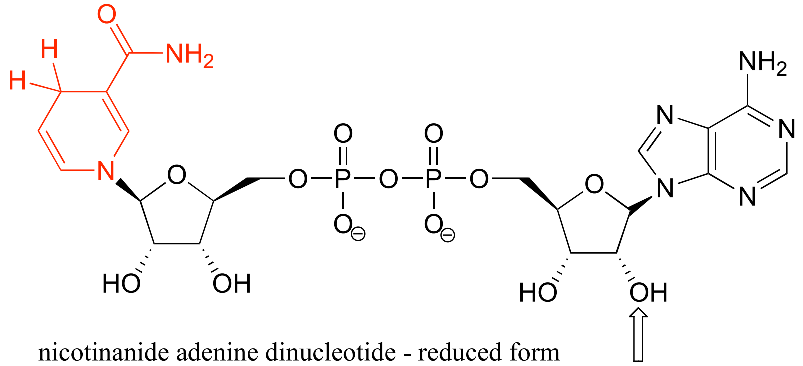
Glutathione
(GSH)



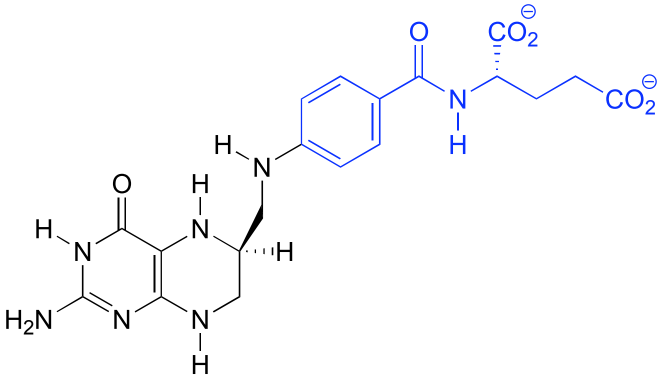
biotin



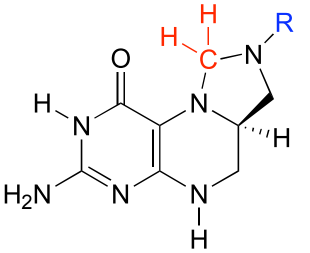
nicotinanide adenine dinucleotide - oxidized form
(NAD$^+$, or NADP$^+$ if phosphorylated at arrow position)



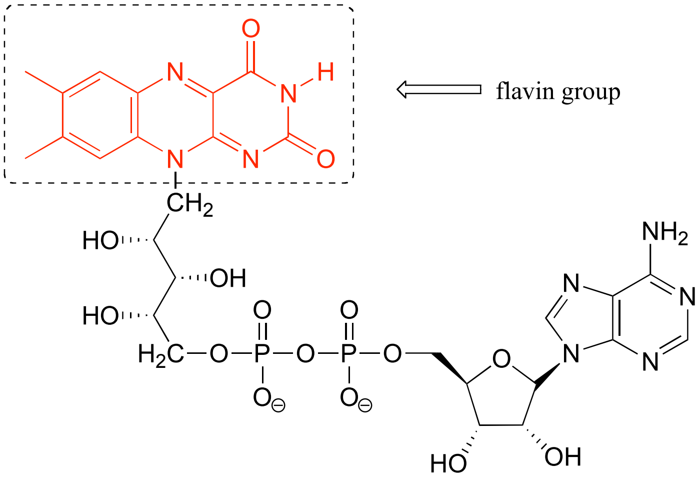
nicotinanide adenine dinucleotide - reduced form
(NADH, or NADPH if phosphorylated at arrow position)



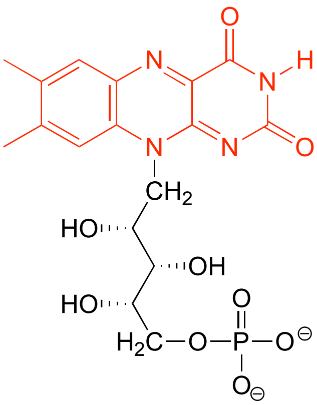
tetrahydrofolate (THF)



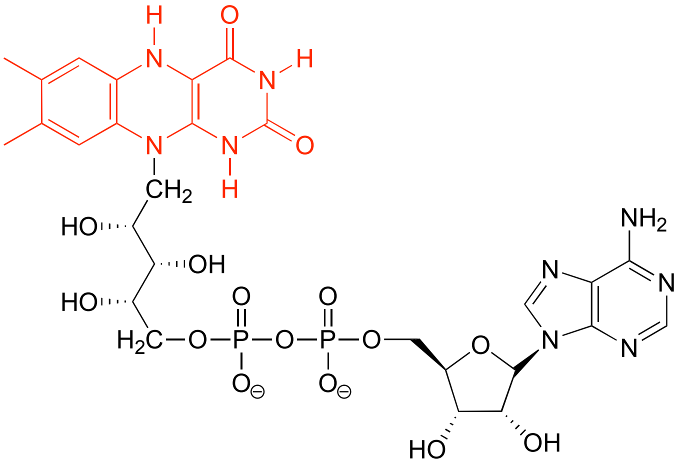
5,10-methylenetetrahydrofolate



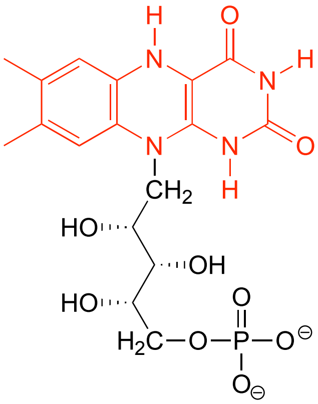
flavin adenine dinucleotide, oxidized form
(FAD)



flavin mononucleotide, oxidized form
(FMN)

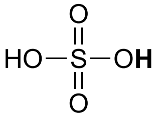
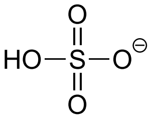
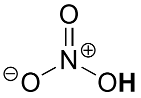
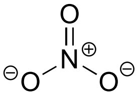
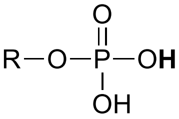
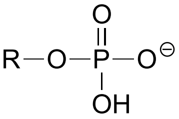
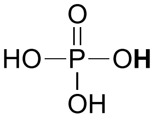
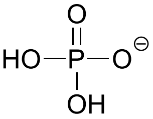
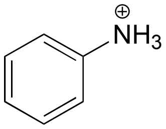
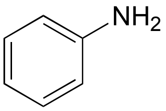
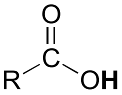
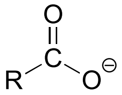
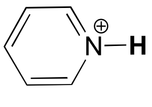
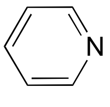


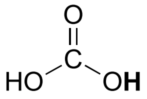
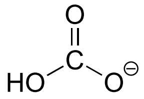
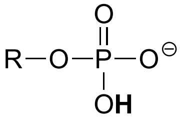
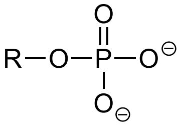
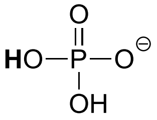
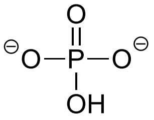
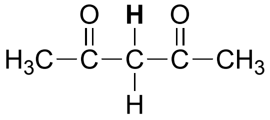
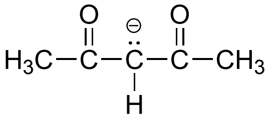
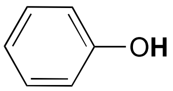
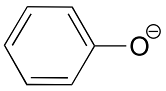
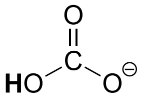
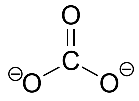
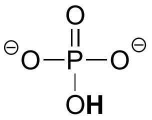
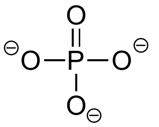
flavin adenine dinucleotide, reduced form
($FADH_2$)

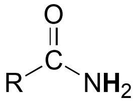
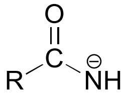
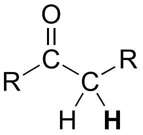
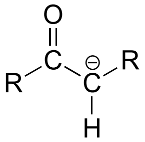


flavin mononucleotide,reduced form
($FMNH_2$)

Table 7: Representative acid constants.

acid	pK_a	conjugate base
 sulfuric acid	-10	
HCl	-7	Cl^-
H_3O^+	-1.7	H_2O
 nitric acid	-1.4	
	$1.0^{(i)}$	
 phosphoric acid	$2.2^{(ii)}$	
HF	3.2	F^-
	4.6	
	4-5	
 pyridinium	$5.3^{(ii)}$	 pyridine

acid	**pK_a**	**conjugate base**
 carbonic acid	6.4	 bicarbonate
	6.5(i)	
	7.2(ii)	
	9.0	
HCN	9.2	CN^-
$\overset{\oplus}{N}\mathbf{H}_4$ ammonium	9.2	NH_3 ammonia
 phenol	9.9(ii)	 phenolate
 bicarbonate	10.3(ii)	 carbonate
$RS\mathbf{H}$	10-11	RS^-
RNH_3^+	10 -11	RNH_2
	12.3(i)	
H_2O	15.7	OH^-

acid	**pK_a**	**conjugate base**
	17	
RCH_2OH	16	RCH_2O^-
	19-20	
$RCC\mathbf{H}$ terminal alkyne	25	RCC^-
H_2	35	H^-
NH_3 ammonia	38	NH_2^-

All pK_a values, unless otherwise noted, are taken from March, Jerry, Advanced Organic Chemistry, Fourth Edition, Wiley, New York, 1992.

(i) Silva, J.J.R. Fraústo da, The Biological Chemistry of the Elements: the Inorganic Chemistry of Life, 2nd Edition, Oxford, New York, 2001.

(ii) Lide, David R. (ed.) The CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, FL, 1995.

Table 8: Some common laboratory solvents, acids, and bases

Solvents

CH_3CH_2OH

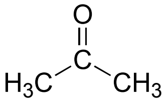
ethanol

CH_3OH

methanol

H_2O

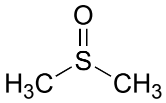
water



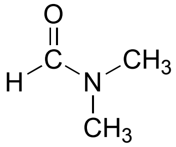
acetone

$N{\equiv}C{-}CH_3$

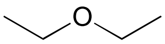
acetonitrile



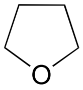
dimethyl sulfoxide
(DMSO)



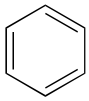
dimethylformamide



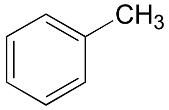
diethylether



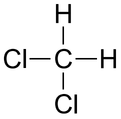
tetrahydrofuran
(THF)



benzene

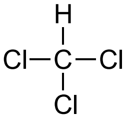


toluene



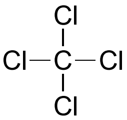
$$\mathrm{Cl-\overset{\displaystyle H}{\underset{\displaystyle Cl}{C}}-H}$$

dichloromethane
(methylene chloride)



$$\mathrm{Cl-\overset{\displaystyle H}{\underset{\displaystyle Cl}{C}}-Cl}$$

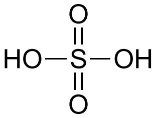
chloroform



$$\mathrm{Cl-\overset{\displaystyle Cl}{\underset{\displaystyle Cl}{C}}-Cl}$$

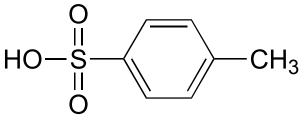
carbon tetrachloride

Acids

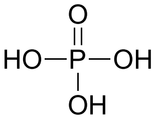


$$\mathrm{HO-\overset{\displaystyle O}{\underset{\displaystyle O}{S}}-OH}$$

sulfuric acid



p-toluenesulfonic acid
(TsOH)



$$\mathrm{HO-\overset{\displaystyle O}{\underset{\displaystyle OH}{P}}-OH}$$

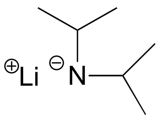
phosphoric acid

HCl

hydrochloric acid

Bases

very strong bases:



lithium diisopropylamide
(LDA)

$^{\oplus}Li\ ^{\ominus}CH_2CH_2CH_2CH_3$

N-butyllithium

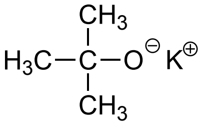
NaH

sodium hydride

$^{\oplus}Na\ ^{\ominus}NH_2$

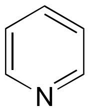
sodium amide

weaker bases:

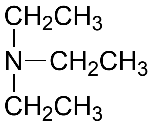


$H_3C-C(CH_3)_2-O^{\ominus}\ K^{\oplus}$

potassium *tert*-butoxide

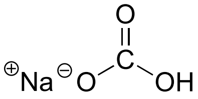


pyridine

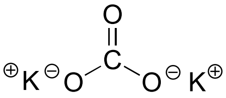


$N(CH_2CH_3)_3$

triethylamine



sodium bicarbonate
$NaHCO_3$

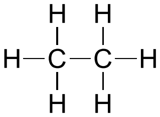
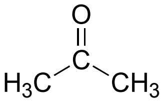
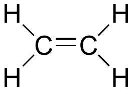
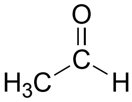
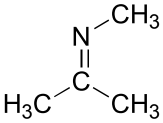
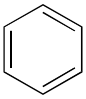
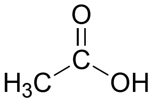
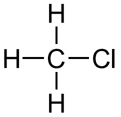
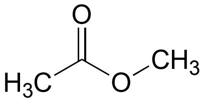
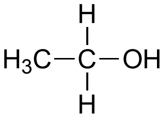
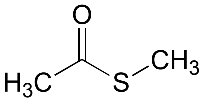
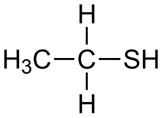
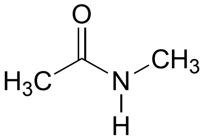
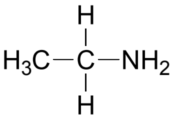
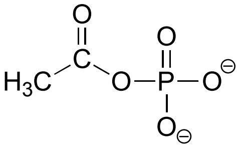
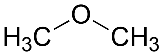
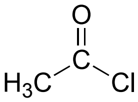
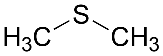
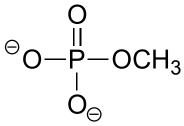
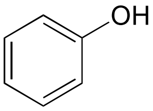
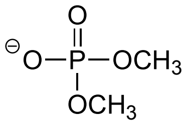


potasium carbonate
K_2CO_3

$NaOH$

sodium hydroxide

Table 9: Examples of common functional groups in organic chemistry

alkane	 $H-\underset{H}{\overset{H}{C}}-\underset{H}{\overset{H}{C}}-H$	**ketone**	
alkene		**aldehyde**	
alkyne	$H-C\equiv C-H$	**imine (Schiff base)**	
aromatic hydrocarbon		**carboxylic acid**	
alkyl halide	 $H-\underset{H}{\overset{H}{C}}-Cl$	**ester**	
alcohol	 $H_3C-\underset{H}{\overset{H}{C}}-OH$	**thioester**	
thiol	 $H_3C-\underset{H}{\overset{H}{C}}-SH$	**amide**	
amine	 $H_3C-\underset{H}{\overset{H}{C}}-NH_2$	**acyl phosphate**	
ether		**acid chloride**	
sulfide		**phosphate ester**	 $^{\ominus}O-\underset{O^{\ominus}}{\overset{O}{\overset{\|}{P}}}-OCH_3$
phenol		**phosphate diester**	 $^{\ominus}O-\underset{OCH_3}{\overset{O}{\overset{\|}{P}}}-OCH_3$